DRAFT
9-29-99

GUIDANCE FOR EVALUATING RESULTS OF AQUATIC TOXICITY TESTS

The purpose of the EVISTRA (EValuation and Interpretation of Suitable Test Results in AQUIRE) database is to present results that (a) were obtained from aquatic toxicity tests on selected chemicals and (b) were evaluated for suitability and quality and interpreted when necessary. (Results of aquatic toxicity tests are neither evaluated nor interpreted before they are entered into the AQUIRE database. The "Documentation Code" in AQUIRE concerns the quality of the report, not the quality of the test result.)

1. The purpose of the suitability evaluation is to reject test results that cannot be used in the derivation of aquatic life benchmarks (i.e., numeric descriptors of the toxicities of chemicals to aquatic life, such as national and site-specific aquatic life criteria, species mean acute values, etc.).
2. The purpose of the quality evaluation is to assess the level of confidence that a similar test result would be obtained from a very high quality test in a different aquatic toxicology laboratory using the same test species and life stage, comparable dilution water, test conditions, etc. Quality evaluations are performed only on suitable test results and take into account how much information is available, and what the information says, concerning (a) the method used for the toxicity test (i.e., the procedures, organisms, test material, dilution water, etc.), and (b) the data obtained concerning the test conditions, concentration-effect curve, etc. (A separate consideration that could affect the level of confidence is whether the test result is supported by results of other toxicity tests, but this cannot be considered when each test result is evaluated based on its own merits. This can, however, be taken into account when the level of confidence in a mean value for a taxon is determined.)

Results entered into EVISTRA are from tests of the toxicity of a single test material to an aquatic taxon through a water-column exposure. The entered results are raw toxicity data when they are available; otherwise, the entered results are calculated endpoints. Supplementary information (e.g., results of quality evaluations, taxonomic and geographic information concerning test species, etc.) and software can be accessed to facilitate use of test results in EVISTRA. This guidance first discusses checklists and suitability and quality evaluations and then presents suitability and quality checklists, and instructions for using each checklist.

This guidance was written by Charles Stephan, with much help from Dean Cox, Steve Broderius, Tim Dawson, Greg Elonen, Russ Erickson, Jim McKim, Dave Mount, and Chris Russom. Comments and questions concerning this guidance, including the checklists and instructions, should be sent to Charles Stephan, U.S. EPA, 6201 Congdon Blvd., Duluth, MN 55804 (TEL: 218-529-5219; FAX: 218-529-5003; EMAIL: stephan.charles@epa.gov).

Checklists

To reduce subjectivity, both suitability and quality checklists are used in the evaluation of results of toxicity tests with aquatic organisms. Because of important differences between various types of aquatic toxicity tests, both kinds of checklists must be individually designed for results of acute tests, early life-stage tests, etc. Most of the items on the checklists are presented as questions such that a negative answer to the question indicates a problem concerning the test result. For each checklist, a set of instructions gives explanatory information concerning items on the checklist. Although the checklists and instructions address many details in order to maximize uniformity from one reviewer to another, numerous judgments will be necessary. Therefore, the checklists are intended to be used by people who have conducted aquatic toxicity tests. (Note: This version of the draft guidance only contains checklists for acute toxicity tests; other types of toxicity tests will be addressed later.)

Suitability Evaluation

After EPA selects a chemical, each test result that might be useful in the derivation of a benchmark for that chemical is first evaluated using a suitability checklist, and each suitable result is then evaluated using a quality checklist. There are several important differences between the two evaluations:

A. The purpose of the suitability evaluation is to reject test results that cannot be used in the derivation of aquatic life benchmarks, whereas the purpose of the quality evaluation is to assess the level of confidence that can be placed in a suitable test result.
B. All of the issues addressed in the suitability evaluation are expressed as “pass-fail” issues, whereas the quality evaluation addresses as few “pass-fail” issues as is practical.
C. “Unknown” is an acceptable answer to issues addressed in the quality evaluation, but not in the suitability evaluation.

It is important that the suitability evaluation reject test results that cannot be used in the derivation of aquatic life benchmarks because it is a waste of resources to evaluate the quality of unsuitable test results. However, the suitability evaluation should not reject a test result that might be useful in some situations, even if it is not likely to be useful in very many situations, because EVISTRA is intended to be useful in the derivation of aquatic life benchmarks that are applicable to a variety of situations. For example:

a. The pH of the dilution water might be too high for some situations, but acceptable for other situations.
b. In field situations, the tested life stage of the test species might occur only rarely in the type of water (fresh or salt) used in a particular test.
c. In field situations, the tested life stage of the test species might occur only rarely at the temperature used in a particular test.
d. A taxon might be relevant to some, but not all, geographic areas.

e. A taxon might be relevant to some, but not all, habitats. For example, although the following species live in uncommon habitats, data concerning their sensitivities might be useful in some situations:
    1. The pitcher plant mosquito, *Wyeomyia smithi*, lives only in small bog pools and in water that has accumulated in pitcher plants (Pennak 1978). [The name is spelled *Wyeomyia smithii* in Merritt and Cummins (1996).]
    2. Brine shrimp (i.e., species in the genus *Artemia*) usually only occur naturally in water with a salinity greater than 35 g/kg.
    3. *Caecidotea bicrenata* is a blind cave isopod (Bosnak and Morgan 1981).

The EVISTRA database contains information and is structured to facilitate identification of results that are useful in some situations but not in others.

Quality Evaluation

If a test result passes the suitability evaluation, it is then evaluated using a quality checklist. In order to be an appropriate basis for a determination of the level of confidence in a test result, the quality checklist needs to take into account all aspects of the procedures used and data obtained that relate to the level of confidence that a similar test result would be obtained from a very high quality test in a different aquatic toxicology laboratory using the same test species and life stage, etc. The items addressed by the quality checklist do not necessarily agree with either (a) the information that is and is not usually published in reports of the results of toxicity tests with aquatic organisms, or (b) the recommendations that are contained in various written methods for conducting aquatic toxicity tests. Quality evaluations and checklists such as these might lead to improvements in the methods used to conduct aquatic toxicity tests and to improvements in reports of the results of aquatic toxicity tests.

The following are not used in the evaluation of the quality of a test result even though it might seem that they ought to be useful:

A. Reference toxicant tests are not used because results of such tests using a toxicant whose mode of action is different from the mode of action of the test material can imply a problem when none exists or can imply the absence of a problem when one does exist (McNulty et al. 1999). Additional research concerning modes of action, etc., might increase the usefulness of reference toxicant tests.
B. Confidence limits on such endpoints as LC50s are not used because there are reasons to believe that many confidence limits are not calculated correctly. For example, some methods and programs used to calculate confidence limits do not account for between-chamber variability when a test contains replicate experimental units for one or more treatments. (In particular, some methods and programs used to calculate confidence limits on LC50s, EC50s, etc., do not allow for the possibility of extra-binomial variation.) In addition, the confidence limits obtained using different calculation methods and programs differ much more than the endpoints obtained using different calculation methods and programs. Additional development

of calculation methods and programs might increase the usefulness of confidence limits on endpoints.

C. The source of the dilution water is not used because it is not clear that such information is useful. For example, many river waters are contaminated, but some are not. Similarly, many ground waters are not contaminated, but some are. In addition, the qualities of some river waters and some well waters vary substantially over time. Data concerning the ability of aquatic organisms to survive, grow, and/or reproduce in the dilution water at the time of the toxicity test are more useful for evaluating the quality of the dilution water than information concerning the source of the dilution water, unless waters from different sources were used at different times during a test.

D. Results of chemical analyses of dilution water are not used because it is virtually impossible to (1) analyze for all possible contaminants and (2) adequately interpret the measured concentrations, even without considering additivity, synergism, and antagonism, unless the result of only one chemical analysis is very unusual. Data concerning the ability of aquatic organisms to survive, grow, and/or reproduce in the dilution water at the time of the toxicity test are more useful for evaluating the quality of the dilution water than information concerning contaminants and other chemicals in the dilution water.

E. The source of the test organisms is not used in the evaluation of a test result because of opposing considerations. It can be argued that such things as quality of food, history of parents, conditions before the test, etc., can be addressed better for organisms raised in culture units and hatcheries than for organisms obtained from natural bodies of water. Alternatively, it can be argued that organisms from field situations are more natural and less likely to be affected by inbreeding. Whether all of the test organisms in a test are from the same source affects the rating, but the source itself does not affect the rating.

F. Measurements of the concentration of test material in stock solutions are not used because they do not provide direct information concerning the concentration of test material in test solutions. Many problems can occur in the preparation of test solutions from stock solutions.

G. No use is made of upper or lower limits on pH or temperature because site-specific values of pH and temperature range from very high to very low. It is not realistic to claim that a test is low quality just because it was conducted at, for example, a low pH in order to match the pH at a site. In addition, the values of pH and temperature that stress aquatic organisms are species-specific.

H. An upper limit on Total Suspended Solids (TSS) is not used because concentrations of TSS that are sufficiently high to stress aquatic organisms are likely to occur in dilution water very rarely. If the concentration of TSS is sufficiently high to be of concern, it can be addressed in the last section of the checklist titled "Miscellaneous". Also, results of tests conducted at high concentrations of TSS might be more useful in some situations than results of tests conducted at low concentrations of TSS.

I. The degree of agreement between, for example, an LC50 and the concentration-effect data is not used because it cannot be used. If the concentration-effect data

are available, they are entered into EVISTRA; if the data are not available, the data cannot be entered and the degree of agreement cannot be evaluated.

An important consideration concerning evaluation of the quality of a test result is the distinction between “quality” and “utility” of test results. Although each factor that reduces the quality of a test result also reduces the utility of the test result, there are factors that reduce utility but do not reduce quality. For example, two test results of equal quality might have different utilities in a particular situation because, for example, the pH of one of the dilution waters might have been much closer to the pH of the site water than the pH of the other dilution water. All of the factors that affect confidence in a test result should be considered in the evaluation of quality, whereas factors that affect utility, but not quality, should not be considered in the evaluation of quality.

Deciding whether a factor does or does not reduce the quality of a test result is sometimes difficult, but the following general considerations address most concerns:

1. If a factor reduces confidence in a test result (i.e., reduces confidence that a similar test result would be obtained from a very high quality test in a different aquatic toxicology laboratory), it reduces quality; if it does not reduce confidence, it does not reduce quality.
2. Factors that can reduce the quality of the test organisms can therefore reduce confidence in a test result.
3. Factors that can stress test organisms can therefore reduce the quality of the organisms and can therefore reduce confidence in a test result.
4. A factor that reduces quality should be on the checklist only if it might reasonably occur during toxicity tests. Most rare factors, however, can be adequately covered in general statements. For example, although they are likely to stress test organisms, earthquakes are rare. Earthquakes should not be on the quality checklist, but “disturbance” should be on the list.
5. Some aspects of experimental design affect confidence, but others do not because the experimental design of a toxicity test should take into account not only the desired level of confidence, but also such issues as cost, the possibility of all test concentrations being too high or too low, and the number of test organisms used.
6. A rationale that is specific to a test material or a class of test materials probably concerns utility, not quality. For example, a water quality characteristic (e.g., hardness) that affects the toxicity of a test material or a class of test materials relates to utility and therefore should be recorded in the database in relation to the test result. In contrast, a water quality characteristic should be on the quality checklist only if the value of the water quality characteristic can be sufficiently high, low, or variable during a test that it can stress test organisms.

These general considerations result in the following:

a. Use of unhealthy or unacclimated test organisms reduces confidence in test results.
b. Factors such as disturbance and turbulence can stress aquatic organisms and can therefore reduce confidence.

c. The more differences there are between test chambers, etc., from one treatment to another, the less confidence there is in a test result.
d. Large fluctuations in a water quality characteristic or in the concentration of the test material during a test reduce confidence in the test result more than small fluctuations.
e. As the purity of a test material decreases and the percentage of impurities increases, there is more uncertainty concerning whether the observed effects were caused by the test material itself. If the purity is low, confidence is increased if side-by-side tests have shown that high-purity and low-purity materials with the same active ingredient produce the same results. (Results of such comparisons should usually be extrapolated from one low-purity material to another only if it is likely that the impurities are similar.)
f. Confidence in a test result is reduced by the following:
    1. Concentrations of dissolved oxygen (DO) that are sufficiently low to stress test organisms.
    2. Concentrations of dissolved nitrogen and/or DO that are sufficiently high to stress test organisms.
    3. Fluctuations in pH, hardness (or salinity), and temperature that are sufficiently severe within a test chamber to stress test organisms.

    For such water quality characteristics as temperature and pH, it might seem that it would be of interest to compare the means of the measured values in different test chambers, because it would seem that there would be concern if, for example, the mean measured temperature in one test chamber was substantially different from that in another test chamber. Such comparisons are not likely to be worth the effort, however, for three reasons. First, some tests that have such a problem are also likely to have a problem with, for example, temperature fluctuation within a test chamber, which is covered by item #3 above. Second, randomization of test chambers will minimize the importance of such differences within treatments; non-random assignment of treatments to test chambers reduces confidence in and of itself. Third, in some cases the effect can be taken into account in the analysis and/or use of the data. For example, some test materials can affect the pH of the test solution. Adjustment of pH might be feasible or desirable, but a correlation between pH and the concentration of the test material does not necessarily render the result useless. For some chemicals, it is possible to model a relationship between the speciation of the test material and pH.
g. An absence of irregularities (e.g., inversions) in the data concerning percent effect vs. time and percent effect vs. concentration increases confidence in a test result.
h. There is more confidence in a test result if it is based on measured concentrations of the test material than if it is based on nominal (i.e., intended) concentrations of the test material. Further, the importance of measuring concentrations depends on the properties of the test material and on various aspects of the toxicity test methodology. In addition, for some chemicals, issues concerning test methodology and/or analytical methodology can affect both quality and utility. For example, for such chemicals as chlorine and tetrachlorodibenzo-*p*-dioxin (TCDD), the absence of measurements or the use of unacceptable analytical methods can make the test

results unsuitable for use in the derivation of benchmarks. Similarly, although both total recoverable and dissolved methods are acceptable for measuring the concentrations of metals in test solutions, many scientists believe that the toxicity of a metal in the water column correlates better with the concentration of dissolved metal than with the concentration of total recoverable metal. Because such issues can have a large effect on utility, it is important that they be addressed on a chemical-specific basis during the suitability evaluation.

Entries on Quality Checklists

When a test result is reviewed using a quality checklist, an entry must be made for each item on the checklist. This ensures that each item is considered and helps reviewers keep track of which items have and have not been addressed. Only four entries may be used:

1. A dash indicates either "Yes" or "Not Applicable".
   It does not make any difference which one a dash indicates because "Yes" is positive and "Not Applicable" is neutral, whereas the purpose of the checklist is to track entries that indicate problems.
2. N indicates "No" in response to a question and is always a negative entry.
   Ns are so important that each one should be based on information which indicates that there is a problem concerning a test result so that obtaining additional information regarding the test result will rarely reduce the number of Ns. Therefore, items need to be worded so that Ns are not entered because of a lack of information. Each item and the instructions should be worded so that the entry is U, not N, if sufficient information is not available. Nevertheless, if unnecessary Ns are assigned to test results because documents contain unclear, misleading, or incorrect information, new information might reduce the number of Ns.
3. U indicates "Unknown" and is a moderately negative entry.
   U indicates that sufficient information is not available and that it is not possible to infer an entry from another document by the same author(s) or by means of a reference to a written method (see below). If sufficient information was available, the entry would be a dash, N, or M. When EPA wants tests evaluated for EVISTRA, assumptions should not be made unless specified in the instructions or approved by the EVISTRA coordinator. If no relevant information is available, the entry will usually be U. Because U indicates "unknown", it is not as negative as N. (It might be thought that some items are so important that not having the relevant information should cause the entry to be N because the lack of the information reduces confidence more than is indicated by U. In such cases, it is better to assign more Us than to assign even one N.)
4. M indicates "referenced to a written Method" and is a weakly negative entry.
   M has a negative connotation because information is not explicitly reported concerning the item, but is implied by reference to an available written method. For the purposes of this guidance, a "written method" is a method that was

developed by a committee or organization, or an individual other than one of the investigators who conducted the test being reviewed. Methodology published in another document by the same author(s) is not considered a "written method" because it seems reasonable to assume that authors would note all deviations from their own methods. To minimize misunderstandings, the reviewer must read the written method to verify what it says.

Reference to a written method is weak for two reasons:

a. Some written methods contain both mandatory (e.g., "must") and optional (e.g., "should") guidance and it is valid to say that the method was followed even if none of the optional guidance was followed.
b. Some investigators report that a particular written method was followed, even though they did not follow the method regarding one or more aspects of the test. Thus, whenever a written method is referenced, it is not possible to know what was actually done, unless explicit information is available.

Because a reference to a written method is better than no information concerning an aspect of a test, M is not as negative as N or U. If a document not only refers to a written method but also lists specific exceptions to the written method, the reference to the written method is considered to provide explicit information; therefore, a dash or N, rather than M, is to be entered.

The wordings of the items on the quality checklists are such that each N, U, and M indicates less confidence in the result of a toxicity test than a dash.

If an appropriate person can be contacted, a questionnaire is sent to request the additional information that is needed to reduce the number of Us and Ms assigned to a test result.

Six-Level Rating Scheme

In existing guidance (U.S. EPA 1985a), two different kinds of schemes are used to address the quality of results of toxicity tests:

A. Pass-Fail Scheme.

Each test result is either used or unused. Some unused results, however, are not necessarily incorrect; they are merely questionable. For example, if the test organisms were not acclimated to the test conditions before the beginning of an acute test, the result is not used even if it is not actually known whether a test result obtained with a group of organisms that had been acclimated would be different from a test result obtained with a group of the same organisms that had not been acclimated.

B. Prioritization Scheme.

For most test materials, "flow-through, measured" acute values are given priority over "static, measured" acute values, even though both are considered acceptable. Thus, whether the result of a "static, measured" acute test is used in

calculations depends partly on whether the result of a comparable “flow-through, measured” acute test is available.

Because the above schemes do not adequately address the quality of a test result, a multilevel quality rating scheme was developed to express the level of confidence in a test result, i.e., the level of confidence that a similar test result would be obtained from a very high quality test in a different aquatic toxicology laboratory using the same test species and life stage; comparable test material, dilution water, and test conditions; and acceptable procedures to conduct the test and calculate the result. This rating scheme has the following six levels:

1. Very high quality.
   There is a very high level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory, although there is a slight possibility that a similar test result would not be obtained. A substantial amount of information is available concerning the procedures and the data, and, although there might be a few very small problems, there is no indication of a more serious problem.
2. High quality.
   There is a high level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory, although there is a small possibility that a similar test result would not be obtained. There is a small problem concerning the procedures and/or data and some desirable information concerning the procedures and/or data might not be available.
3. Moderate quality.
   There is a moderate level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory, although there is also a moderate possibility that a similar test result would not be obtained. There is a medium problem or more than one small problem concerning the procedures and/or data and some desirable information concerning the procedures and/or data might not be available.
4. Low quality.
   There is a low level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory, although there is a small possibility that a similar test result would be obtained. There is a large problem or more than one medium problem and/or several small problems concerning the procedures and/or data and some desirable information concerning the procedures and/or data might not be available.
5. Very low quality.
   There is a very low level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory, although there is a slight possibility that a similar test result would be obtained. There is a very large problem and/or more than one large problem concerning the procedures and/or data and some desirable information concerning the procedures and/or data might not be available.
6. Unknown quality.

Sufficient information is not available to assess the level of confidence in the test result.

Weights on Quality Checklists

In order for a quality checklist to be meaningful, each assigned N needs to represent a comparable amount of reduction in confidence. Therefore, because some items are more important than others, it is necessary to assign weights to the items, with the least important items having a weight of 1.

A. If the weight for the item is always the same, the weight is given on the checklist. These weights apply not only to N, but also to U and M.
B. If the weight for the item depends on the value of a variable concerning the test result, such as the percent purity of the test material, a weight is not given on the checklist and it is necessary to read the instructions to determine the weight to assign. When possible, a sliding scale is used to determine the weight; otherwise, a stepwise scale is used. Each sliding scale is expressed as an equation that relates the number of Ns to the value of the variable, except that the maximum weight allowed for any item is 40. When a sliding scale is used, the weight to be used with an entry of U or M is specified. When a stepwise scale is used, the weight to be used with an entry of U or M is usually the highest weight in the scale.

A variety of factors (e.g., exposure to unusual chemicals, fluctuating temperature, turbulence, etc.) can reduce confidence because they can stress test organisms, but the sensitivities of organisms to such factors are probably species-specific. Because species-specific weights for such items are not feasible, general weights are used.

The most rational way to assign weights to the items is to relate the weights to an interpretation of the number of Ns. Therefore, when each entry on a quality checklist is either a dash or N, the total number of Ns is assigned one of the following quality ratings:

Number of Ns	Quality Rating
0 to <10	Very high
10 to <20	High
20 to <30	Moderate
30 to <40	Low
≥40	Very low

The upper limit of <10N for “very high quality” was selected by considering how many items with a weight of 1 could be assigned N and still consider that there is a very high level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory. A test result with up to 10 more Ns would be “high quality”, and each increment of 10 more Ns would reduce the quality rating so that a test result with at least 40Ns would be “very low quality”.

The least important items were assigned a weight of 1 and slightly more important items were assigned weights less than 10. Each item for which N on that item alone would drop the rating to "high quality" was assigned a weight of 10. Thus, N on two such items would drop the rating to "moderate quality" and N on four or more such items would drop the rating to "very low quality". In addition, each item for which N on that item alone would drop the rating to "very low quality" was assigned a weight of 40. The weights for other items were assigned based on the relative importance of the items.

The assignment of weights can be illustrated using examples related to measurement of the concentrations of the test material (see item h on page 6).

a. If an individual aspect of the measurement process was not performed as desired during a test on a test material such as sodium chloride, this is a very small problem and therefore the weight is <10N. If this is the only problem, such a test result is still very high quality because there is a very high level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory.
b. If no measurements were performed during a test on a test material such as sodium chloride, this is a small problem and therefore the weight is 10N.
c. If no measurements were performed during a test on a test material such a copper chloride, this is a medium problem and the weight is 20N.
d. If no measurements were performed during a test on a test material such as chlorine, this is a large problem and the weight is 30N.
e. If no measurements were performed during a test on a test material such as TCDD, this is a very large problem and the weight is 40N. A result from such a test would be very low quality because there is a very low level of confidence that a similar test result would be obtained from a very high quality test in a different laboratory.

As discussed in item h on page 6, for such chemicals as chlorine and TCDD, issues concerning test methodology and measurement of the concentration of the test material in test solutions can affect both quality and utility. For such chemicals, the absence of measurements or use of unacceptable analytical methods can severely reduce the level of confidence that a similar result would be obtained in a different laboratory and can make the test results unsuitable for use in the derivation of benchmarks.

Summary Scores and Quality Ratings

The above relationship between the weights and the rating scheme works only when each entry is a dash or N, because even a result that is assigned no Ns cannot be considered "very high quality" if it has too many Us and/or Ms. There are two situations:

1. A test result with fewer than 10Ns should be "unknown quality" if it has an excessive number of Us and/or Ms, and this should not depend on the number of Ns. Because U expresses more uncertainty than M, it seems reasonable to calculate an "Uncertainty Score" by assigning 3 points for each U and 1 point for each M. Therefore: Uncertainty Score = (3)(# of Us) + (# of Ms).
2. A test result with fewer than 10Ns should be rated "high quality", rather than "very high quality", if it has an intermediate number of Us and/or Ms. In addition, a test

result with 2Ns should need more Us and/or Ms than a test result with 5Ns to be downgraded to "high quality". Because N is more negative than U, and U is more negative than M, it seems reasonable to calculate a "Negativity Score" by assigning 10 points for each N, 3 points for each U, and 1 point for each M. Therefore: Negativity Score = (10)(# of Ns) + (3)(# of Us) + (# of Ms).

The same rationale applies when the number of Ns is between 10 and 20 or between 20 and 30. In contrast, a test result with 30 or more Ns should be rated either "low quality" or "very low quality" regardless of the Uncertainty Score.

To appropriately take into account the numbers of Ns, Us, and Ms, quality ratings are assigned as follows:

Number of Ns	Uncertainty Score	Negativity Score	Quality Rating
0 to <10	<100	<100	Very high
0 to <10	<100	≥100	High
0 to <10	≥100	*	Unknown
10 to <20	<100	<200	High
10 to <20	<100	≥200	Moderate
10 to <20	≥100	*	Unknown
20 to <30	<100	<300	Moderate
20 to <30	<100	≥300	Low
20 to <30	≥100	*	Unknown
30 to <40	*	<400	Low
30 to <40	*	≥400	Very low
≥40	*	≥400	Very low

* = This score is irrelevant in this situation.

The Uncertainty Score weights and integrates the numbers of Us and Ms, whereas the Negativity Score weights and integrates the numbers of Ns, Us, and Ms. The limits of 100, 200, and 300 for the Negativity Score and the limit of 100 for the Uncertainty Score are based on the rationale that (a) 10, 20, and 30Ns are required for a test result to be downgraded to "high quality", "moderate quality", and "low quality", respectively, when each entry is either a dash or N, and (b) the relative weights are 10, 3, and 1 for N, U, and M, respectively. Thus, a test result with zero Ns and zero Ms can have 33Us and still be "very high quality". Similarly, a test result with zero Ns and zero Us can have 99Ms and still be "very high quality".

When quality evaluations are performed for EVISTRA, the whole process is computerized. For many of the items on the checklist, the reviewer enters the relevant information and the program performs the necessary calculations. The numbers of Ns, Us, and Ms assigned to each test result are available from the database so that users can see these numbers if desired. Software assigns a quality rating to each test result using the six-level rating scheme described above.

This quality rating scheme is a nonstatistical way of integrating a variety of considerations into a single rating. The rating assigned to a test result using this scheme is based on expert judgment. There is no statistical basis for the quality ratings and no statistical implications are implied; therefore, no percentages can be associated with the quality scores and ratings. It might be possible to evaluate a sufficiently large dataset and develop a statistically based rating scheme, but this has not been done and is not planned.

Scores vs. Ratings

It is useful to translate Negativity Scores and Uncertainty Scores into Quality Ratings because the ratings give meaning to the scores and provide a basis for assigning weights. Nevertheless, problems can arise when Uncertainty Scores and Negativity Scores are converted to Quality Ratings because the scores consist of many small steps whereas the ratings consist of a few large steps. For example, when quality ratings are based on limits such as 100, 200, and 300, two test results can be assigned different ratings even though the scores assigned to these test results differ by only one point in one of the summary scores. This also means that rounding can be important. Because some test results will have borderline scores, it is important that appropriate rules be applied consistently when weights are calculated, rounded, and summed. More importantly, judgment is used in many places in the quality checklist (e.g., in the interpretation of vague information) and so it is likely that even trained reviewers will not always assign the same Negativity Score or Uncertainty Score to a test result. Therefore, whenever possible, it is probably desirable to use the Negativity Score and/or the Uncertainty Score, rather than the Quality Rating, to avoid the potential problems that can result from the large steps. For example, if species mean acute values are calculated it might be possible to weight each test result according to the reciprocal of the Negativity Score and/or Uncertainty Score instead of using an accept-reject system based on the ratings or scores.

Additional Material

Books by Fava et al. (1994) and Swanson and Socka (1997) provide some interesting information concerning topics discussed above and related issues, but neither is directly relevant. Fava et al. (1994) address assessment of datasets much more than assessment of individual data whereas Swanson and Socka (1997) apply ranking and

scoring to the assessment of chemicals on the basis of hazard, persistence, and exposure. Both of these books and Finkel (1990) comment on the importance of addressing uncertainty, but none describes a way of addressing uncertainty that is appropriate for the process described herein for the evaluation of the quality of an individual test result. Seiler and Alvarez (1998) provide some interesting comments regarding kinds of uncertainty and the importance of the standard for expressing uncertainty from the American National Standards Institute.

9-29-99

ACUTE SUITABILITY CHECKLIST

This checklist should not be used unless the "Instructions for Using the Acute Suitability Checklist" have been read, especially the general instructions.

First author: _____ RETOX #: _____
Test organism (scientific name, age/size): _____
Chemical(s): _____ Test(s): _____
Reviewer:_____ Date:_____

Enter either: Dash = Yes or N = No

1 Enter N if there is a chemical-specific reason to reject the test result. . . . . __
2 Is the document written in English? . . . . . __
3 Are the author(s) and date of the document given? . . . . . __
4 Is the document available for public distribution? . . . . . __
5 Is the document a primary source for the test result? . . . . . __
6 Was the stressor a single test material? . . . . . __
7 Is the test material appropriate? . . . . . __
8 If a solvent was used to prepare the test solutions, was it a solvent that is miscible with water? . . . . . __
9 Is the test species in the kingdom Animalia? . . . . . __
10 Were the test organisms whole living organisms? . . . . . __
11 Were the test organisms sufficiently identified and taxonomically similar? . . . . . __
12 Were the duration and adverse effect appropriate for the test organisms? . . . . . __
13 Enter N if any of the test organisms in any important treatment were in a nonaquatic life stage at any time during the test, unless the definition of the adverse effect included effects on the transition to the non-aquatic life stage. . . . . __
14 Was the test result based only on test organisms that were exposed to the treatments from the beginning of the test? . . . . . __
15 Were the test organisms in each important treatment continuously exposed to a consistent concentration of test material throughout the test? . . . . . __
16 Were the test organisms substantially exposed to the test material only via the test solutions? . . . . . __
17 Enter N if there is an organism-specific reason to reject the test result. . . . . __
18 Enter N if the data are from a mesocosm or model ecosystem test. . . . . __
19 Did the dilution water contain sufficient ions that it was not an "ion-free" water? . __
20 If the dilution water was prepared from chlorinated water, is there an explicit reason to believe that the concentration of chlorine was acceptably low? . . . . . __

Comments:_____

INSTRUCTIONS FOR USING THE ACUTE SUITABILITY CHECKLIST

A set of general instructions is given, and then specific instructions are given concerning individual items on the checklist.

General Instructions

a. An N should be based on information that is known about a test result; an N should not be based on a lack of information.
b. If a test result is known to be unsuitable, it is only necessary to address one relevant item on this checklist because a test result will not be entered into EVISTRA if even one entry is N on the suitability checklist. Because only one N is sufficient to cause rejection, it is not necessary to address the items in order or to address all of the items.
c. This checklist should be applied to acute toxicity test results individually, except that it may be applied to two or more acute test results in the same document if (a) the same entry is N for the two or more test results or (b) all of the entries are dashes for the two or more test results.
d. Information in another document by the same author(s) may be used in a suitability evaluation of an acute test result, but the other document must be read to verify what it says. In addition, the other document must be cited as a comment at the end of the checklist even if it is cited in the document being reviewed.
e. Save this checklist only if one or more entries are N or if another document is cited at the end of the checklist. It is not necessary to save this checklist if the entry for each item is a dash because an acute test result must not receive any Ns in order for it to be evaluated using the Acute Quality Checklist.
f. If the test results are otherwise suitable, results reported as “less than” or “greater than” values and results of exposures to concentrations that are above the solubility limit of the test material in the dilution water are suitable. For example, automatic rejection of results of exposures to concentrations that are above the solubility limit might unnecessarily reject useful data concerning resistant test organisms.

Specific Instructions

Item 1

Some test results cannot be used in the derivation of aquatic life benchmarks because the test methodology and/or analytical method used was not appropriate for the chemical. Examples of such potential problems are:

a. Static tests on some chemicals are acceptable, but static tests on such chemicals as chlorine and TCDD that will volatilize, hydrolyze, biodegrade, or sorb to a great extent during a test might not be acceptable.

b. Results of tests on some chemicals might be suitable only if the concentration of the "toxic species" can be measured or estimated acceptably. For example, results of tests on chlorine might be acceptable only if an acceptable analytical method was used. Also, results of tests on TCDD might be acceptable only if the concentration of freely dissolved TCDD can be estimated. Estimation of the concentration of the "toxic species" might be quite difficult, however, if even one reaction rate has to be taken into account.
c. Feeding organisms might be acceptable during tests on some chemicals, but not others. In some cases, it might depend on the analytical method used.
d. Transformation of the test material during tests on such chemicals as iron(II) might cause the test organisms to be simultaneously exposed to two toxicants.

Before suitability evaluations are begun for a chemical selected by EPA, such potential problems must be addressed on a chemical-specific basis by the EVISTRA coordinator if the test results are being evaluated for EVISTRA.

Item 2

This is a requirement merely because most people working on EVISTRA and most people who derive aquatic life benchmarks in the U.S. are not sufficiently fluent in any other language. An English translation is suitable because it is written in English. If a translation is used, a copy of the original document in the original language must be attached to the translation.

Item 3

A document may be a publication, manuscript, letter, thesis, report, memorandum, etc. A test result does not have to be published or peer reviewed to be reproducible in a different laboratory and suitable for the EVISTRA database. Although a test result might be reproducible even if the author(s) and date of the document are not known, a test result is not considered suitable for EVISTRA unless the author(s) and date are known because it is difficult to have much confidence in a test result if it cannot be attributed to a specific person. (If a person is not cited but an organization is cited, the EVISTRA staff will attempt to contact the organization to try to obtain the name of the person who ran the test and/or the name of the principal investigator.)

Item 4

Enter N if the document has "confidential" or similar words on it; otherwise enter a dash. A test result cannot be used in the derivation of benchmarks if it is not available for public distribution.

Item 5

Only primary sources are used in order to minimize problems with transcription errors. Except as noted below, a document is considered a primary source only if at least one of the investigators who conducted the toxicity test is an author.

a. Although the Manual of Acute Toxicity: Interpretation and Data Base for 410 Chemicals and 66 Species of Freshwater Animals (Mayer and Ellersieck 1986)

might be considered a secondary publication in some situations, it is considered a primary publication for the purposes of the EVISTRA database. It is a compilation of acute toxicity tests conducted from 1965 to 1984 at the Fish Control Laboratory, Columbia, MO and its field stations. The original data sheets were reviewed, data quality was considered, and test results were calculated. This resulted in some acute values being slightly different from previously published results. Because the original data sheets were used to verify or recalculate the results, Mayer and Ellersieck (1986) is considered a primary publication.

1. Results of other freshwater acute toxicity tests conducted at the Columbia laboratory or its field stations from 1965 to 1984 should be judged on their own merits, but it must be clear that the results are not in Mayer and Ellersieck (1986) in order to avoid using the same test result twice. It must be remembered that some tests were not included in Mayer and Ellersieck (1986) because the tests were not considered of acceptable quality, but other tests were not included because sufficient documentation could not be found.
2. Mayer and Ellersieck (1986) supersedes Johnson and Finley (1980).

b. Similar considerations apply to the Acute Toxicity Handbook of Chemicals to Estuarine Organisms (Mayer 1987).

Item 6

Benchmarks cannot be derived on the basis of results of tests in which the treatments consisted of co-varying levels of two or more stressors, whether it involved two or more chemical stressors or a combination of chemical and nonchemical stressors. A study concerning additivity, potentiation, or antagonism can be entered into EVISTRA if the study can be treated as two or more toxicity tests that are individually entered into EVISTRA.

Effluents, leachates, drilling muds, fly ashes, sediments, and sludges are not considered single chemicals. In addition, the single chemical cannot be introduced as a component of an effluent, mixture, etc. Emulsifiable concentrates and wettable powders are considered single chemicals because they can provide very high quality results in some cases, which will be determined using the Acute Quality Checklist. Chemicals such as toxaphene are considered single chemicals because they are mixtures of structurally similar organic chemicals that only exist in large quantities as commercial mixtures of the various chemicals and apparently have similar biological, chemical, physical, and toxicological properties. In contrast, the following are not considered single chemicals:

a. Phenol and 2,4-dichlorophenol.
b. Sodium chloride and sodium sulfate.

Item 7

Because the purpose of EVISTRA is to provide test results that can be used in the derivation of aquatic life benchmarks for selected chemicals, some test results are not suitable because the test material used in the toxicity test does not provide

useful data concerning the toxicity of the selected chemical to aquatic life. For example, even if the results of toxicity tests on such test materials as copper sulfide and the EDTA salt of copper can be reproduced in a variety of different laboratories, they are not useful for deriving benchmarks for copper. Therefore, results of tests on such test materials are not entered into EVISTRA. Instructions concerning the appropriateness of the test material cannot be generic, because, for example, instructions that are appropriate for copper might not be appropriate for tributyltin.

Most heavy metals

Salts used must be highly soluble and dissociate completely in water. Therefore, salts containing such constituents as ammonia, cyanide, sulfide, or an organic ion (e.g., acetate, stearate, EDTA) are unacceptable. The following salts are acceptable and unacceptable, as listed.

Metal	Acceptable	Unacceptable
Aluminum	chloride, potassium sulfate, sulfate	acetate, ammonium sulfate
Cadmium	chloride, nitrate, sulfate	acetate, oxide, stearate, sulfide
Chromium(III)	chloride, nitrate, sulfate, potassium sulfate	acetate, oxide, sulfide
Copper	chloride, nitrate, sulfate	acetate, ammonia, oxide, pyrophosphate, sulfide
Lead(II)	chloride, nitrate	acetate, arsenate, fluoroborate, oxide, sulfide
Mercury(II)	chloride, nitrate, sulfate	acetate, cyanide, oxide, sulfide
Nickel	chloride, nitrate, sulfate	acetate, ammonium sulfate, cyanide, oxide, sulfide
Silver	nitrate	acetate, chloride, oxide, sulfide
Zinc	chloride, nitrate, sulfate	acetate, oxide, phosphate, sulfide

As necessary, the EVISTRA coordinator will provide information concerning other salts and test materials used in tests that EPA wants evaluated for EVISTRA.

Item 8

The major problems that can be caused by use of a solvent that is not miscible with water are (1) the test material might not be in the test solutions and (2) the test organisms might contact undissolved solvent.

A solvent is miscible with water if it can be mixed with water in all ratios without separation into two phases; a solvent is not miscible with water if there is any ratio at which a mixture separates into two phases. Acetone, acetonitrile, dimethylformamide (DMF), ethanol, methanol, and triethylene glycol (TEG) are water-miscible organic solvents that can be used to prepare test solutions containing

organic chemicals. As necessary, the EVISTRA coordinator will provide relevant information concerning additional solvents used in tests that EPA wants evaluated for EVISTRA.

Enter a dash if:
a. no solvent was used.
b. each solvent used was either water or a solvent that is miscible with water.

Enter N if a solvent that is not miscible with water (e.g., hexane, xylene) was used.

Item 9

The Acute Suitability Checklist and the Acute Quality Checklist apply only to species in the kingdom Animalia. They do not apply, for example, to plants.

Item 10

For species in the kingdom Animalia, organs, tissues, single cells, and cell cultures are not considered “whole living organisms”.

Item 11

A test result cannot be used in the derivation of a benchmark unless (a) sufficient information is available concerning the identity of the test organisms and (b) the all of the test organisms are sufficiently taxonomically similar. For all organisms except insects, the test organisms must be identified to genus and all of the organisms in a test must be either the same species or two or more species in the same genus. Because it is difficult to identify some young insects, test results with insects are suitable even if the test organisms were identified only to the order. It is always desirable, of course, that test organisms are identified to the genus and to the species if possible.

Item 12

For the purposes of deriving benchmarks, only the following adverse effects and durations of exposure are considered suitable for acute tests:
a. Tests with daphnids, other cladocerans, midges, and phantom midges must provide data concerning immobilization and/or death that occurs during and/or shortly after exposure to a test material for up to 96 hours.
b. Tests with embryos and larvae of barnacles, bivalve molluscs (clams, mussels, oysters, and scallops), sea urchins, lobsters, crabs, shrimps, and abalones must provide data concerning incomplete development of shells and/or death that occurs during and/or shortly after exposure to a test material for up to 96 hours.
c. Tests with all other freshwater and saltwater animal species and older life stages of barnacles, bivalve molluscs, sea urchins, lobsters, crabs, shrimps, and abalones must provide data concerning loss of equilibrium, immobilization, and/or death that occurs during and/or shortly after exposure to a test material for up to eight days.

An exposure that lasted longer than specified above might provide suitable data concerning acute toxicity if effects were observed and reported for intermediate periods of exposure.

If an operational definition of the adverse effect is given, it should be consistent with at least one of the following:

A. Death -
    1. For daphnids, lack of heartbeat.
        a. Many of the LC50s that have been reported for daphnids should be entered into EVISTRA as EC50s based on immobilization because it is not possible to determine whether daphnids are dead without using a microscope to check for heartbeat.
    2. For all other motile organisms, lack of movement, especially the absence of respiratory movements in fish and shrimp, and lack of reaction to gentle prodding.
        a. Embryos that do not hatch are considered to have died.
        b. Insects that do not emerge are considered to have died.

B. Immobilization -
    lack of movement except for minor spontaneous, random activity of appendages.

C. Loss of equilibrium -
    inability to make coordinated movements and to maintain a normal upright position.

D. Incomplete shell development (of larvae of bivalve molluscs) -
    The unaffected organisms are live larvae with completely developed shells. All unaffected shells contain meat; empty shells, even if they are completely developed, are considered affected organisms. Live larvae with misshapen or otherwise malformed shells are considered unaffected organisms if the shells are completely developed and contain meat. [See section 11.7.4 of ASTM Standard E724 (ASTM 1998a).]

If necessary, the name of the adverse effect should be changed to conform with the above.

Enter N if the adverse effect and/or duration is not suitable.

Item 13

The “important treatments” are the following:

a. If the LC50 or EC50 is a “greater than” value, the important treatments are the highest tested concentration and all of the tested concentrations that were within a factor of two of the highest tested concentration.
b. If the LC50 or EC50 is a “less than” value, the important treatments are the lowest tested concentration and all of the tested concentrations that were within a factor of two of the lowest tested concentration.

c. If a point estimate of the LC50 or EC50 is calculated, the important treatments are all of the tested concentrations that were within a factor of two higher or lower than the LC50 or EC50.

These are the treatments that provide the really important information about the LC50 or EC50. Other treatments can provide confirmatory information, but such information is not as useful as that provided by the "important treatments".

Unless transition to a non-aquatic life stage is included in the definition of the adverse effect, data obtained at a point in time during an acute test are not suitable if any of the test organisms in an important treatment were in a non-aquatic life stage at that time. For example, in a test with insects, data are not suitable if any of the insects were flying at the time when the data were obtained, unless emergence was included in the definition of the endpoint. To avoid discarding data that might be useful, it is acceptable to define a new endpoint and calculate a new test result if the necessary information is available. (It is possible that transition to a nonaquatic life stage might affect the sensitivities of the organisms, but specially designed tests would have to be conducted to determine whether organisms in the transition period have a different sensitivity.)

Item 14

The result of an acute test should be based only on test organisms that were exposed to the treatments from the beginning of the test to ensure that the duration of exposure was the same for all of the organisms used in the calculation of the test result. There are situations in which reproduction during an acute test does not invalidate the test:

A. If reproduction occurred during a test with worms, any "new" worms would have been produced by fission, which means that both the "new" and "old" worms were exposed from the beginning of the test.
B. If reproduction occurred during a test with a species that is not a worm, either (1) the young should have been removed from each test chamber soon after they were first observed, or (2) it should have been possible, at the end of the test, to distinguish the young from the old organisms so that the young were not used in the calculation of the result. Even if the young were not removed and could not be distinguished from the initial organisms at the end of the test, the test provides useful information if all of the new and old organisms survived or if all of the new and old organisms died.

The entry is assigned as follows:

a. Enter a dash if the test organisms were worms. (The possible effect of reproduction on the quality of results obtained with worms will be evaluated in the quality evaluation.)
b. Enter a dash if it is known that reproduction did not occur in any of the test chambers. For example, enter a dash if it was reported that reproduction did not occur or if the test was conducted with organisms that could not reproduce or if the conditions were not suitable for reproduction by the test organisms.

c. Even if reproduction occurred, enter a dash if either (1) the young were removed from each test chamber while it was possible to distinguish the young from the old, or (2) the young were distinguished from the old organisms at the end of the test and were not used in the calculation of the result.
d. Enter a dash if young were produced and left in the test chambers, but the result was based on the fact that all of the new and old organisms survived in all of the treatments so that the result was a “greater than” value because it is known that all of the old organisms survived.
e. Enter a dash if young were produced and left in the test chambers, but the result was based on the fact that all of the new and old organisms died in all of the treatments so that the result was a “less than” value because it is known that all of the old test organisms died.
f. Enter N for all tests not covered by one of the above. For example, enter N for a test with a rotifer if young were produced and used in the calculation of an endpoint.

Item 15

Concentrations that incidentally changed during a test are considered “consistent” for the purposes of this checklist.

Enter a dash if it was a static test.

For all other tests, enter N if any of the following are true:
a. The exposure was intermittent.
b. The concentration was purposely varied during the test.
c. There was a mistake, accident, or malfunction that affected the concentration of the test material by more than a factor of two.

Otherwise, enter a dash.

Item 16

A test result is unsuitable if the test organisms were substantially exposed to the test material via food, gavage, sediment, or injection before and/or during the test.

Item 17

Some test results might have to be rejected because the tested life stage might be severely stressed if a special requirement was not satisfied in an acceptable manner, which might have implications for other aspects of the test methodology. For example, the life stage might normally live in sediment or on a host. Such potential problems must be addressed on an organism-specific basis by the EVISTRA coordinator if the test results are being evaluated for EVISTRA.

Item 18

A test is not necessarily a mesocosm test or a model ecosystem test just because one or more test chambers contained more than one test species. Mesocosm tests

and model ecosystem tests are unsuitable because they are designed for the test organisms to consume food organisms that grow in the test chamber during the test. Thus, the test organisms are likely to be substantially exposed to the test material via food. In addition, an effect of the test material on the food organisms might have an indirect effect on the test organisms.

It is not acceptable for the survival of the test organisms to depend on food that grows in the test chambers, but it might be acceptable for food to be placed into the test chambers, although the presence of food in the test chamber might affect the quality rating, as per the Acute Quality Checklist.

Item 19

Water obtained from a distillation, deionization, or reverse osmosis unit is considered “ion-free” water.

Enter a dash if any of the following are true:

a. The dilution water was obtained from a well, spring, or body of surface water, such as a pond, river, stream, reservoir, etc.
b. An ion-free water was merely used to dilute a water in order to reduce hardness, salinity, etc.
c. For a test with freshwater organisms, calcium, magnesium, sodium, potassium, chloride, sulfate, and carbonate and/or bicarbonate were added to an ion-free water to prepare a water in which the test organisms could survive, grow, and reproduce.
d. For a test with saltwater organisms, either a commercially available sea salt or a defined mixture of chemicals that was formulated to simulate sea salt was used to prepare a water in which the test organisms could survive, grow, and reproduce.

Enter N only if there is reason to believe that the dilution water was an ion-free water.

Item 20

Water that was dechlorinated is not chlorinated water. Water from a municipal water supply is assumed to be chlorinated unless it was dechlorinated. Water from other sources is assumed to be nonchlorinated, unless there is reason to believe that it was chlorinated.

Enter a dash if a chlorinated water was not used in the preparation of the dilution water.

If a chlorinated water was used in the preparation of the dilution water, enter a dash if one or more of the following are true:

a. The dilution water received treatment that would have resulted in dechlorination, such as aeration for at least 24 hours, passing through a reverse osmosis unit, or addition of sulfite or thiosulfate.
b. A daphnid, ceriodaphnid, mysid, or the test species was cultured and held in the water in the laboratory for 30 or more days during the time when the toxicity test was conducted. (A 30-day duration is specified to ensure that the water was acceptable for a sufficiently long time.)
c. Total residual chlorine was measured acceptably at least weekly over a period of 30 or more days during the time when the toxicity test was conducted and each measured concentration was $\leq$11 ug/L in a fresh water or $\leq$7.5 ug/L in a salt water (U.S. EPA 1985b).

Otherwise, enter N.

ACUTE QUALITY CHECKLIST

a. This checklist is intended to be used to evaluate test results that did not receive any Ns on the Acute Suitability Checklist.
b. This checklist should not be used unless “Instructions for Using the Acute Quality Checklist” have been read, especially the general instructions.
c. The maximum weight allowed for any item is 40.

First author: _____ RETOX #: _____
Test organism (scientific name, age/size):_____
Chemical(s): _____ Test(s): _____
Reviewer:_____ Date:_____

A. Test Organisms **Dash = Yes or Not Applicable, N = No, U = Unknown, M = Method**

1 Were all of the test organisms from the same source? . . . . . 1__
2 Is it true that the test organisms were not collected using electroshocking, chemical treatment, and/or gill nets during the last ten days prior to the start of the test? . . . . . 10__
3 Were all of the test organisms in the same life stage and of similar size at the start of the test? . . . . . 1__
4 If the test organisms were cladocerans, were they from a culture containing no ephippial females? . . . . . 1__
5 Did the test organisms live in the laboratory either (a) from birth or hatch to the start of the test, or (b) for at least the last ten days prior to the start of the test? . . . . . 10__
6 Is it true that the test organisms were not subjected to a rapid, substantial change in temperature and/or water quality either (a) from birth or hatch to the start of the test, or (b) for at least the last ten days prior to the start of the test? . . . . . 3__
7 Were the test organisms in the dilution water either (a) from birth or hatch to the start of the test, or (b) for at least the last 48 hours prior to the start of the test? . . . . . __
8 Were the test organisms at the test temperature either (a) from birth or hatch to the start of the test, or (b) for at least the last 48 hours prior to the start of the test? . . . . . 10__
9 Were the test organisms properly handled and not subjected to substantial disturbance during the test and either (a) from birth or hatch to the start of the test, or (b) for at least the last 48 hours prior to the start of the test? . . . . . 1__
10 Was the loading acceptable in each test chamber? . . . . . __
11 Is it true that aggression was not observed in any of the test chambers? . . . . . 1__
12 Is it true that food was not withheld from the test organisms for an excessive number of consecutive days before and during the test? . . . . . 10__
13 Is it true that the test organisms were not exposed to an unusual substance either (a) from birth or hatch to the start of the test, or (b) during the last ten days prior to the start of the test? . . . . . __
14 Is it true that the test organisms were not exposed to an unusual substance, other than the test material, during the test? . . . . . 40__

B. Experimental Design

1 Was there a control treatment that was comparable to the other treatments? . . . . . __
2 If the dilution factor was < 0.7, was there more than one test chamber per treatment? . . . . . 3__
3 Were all of the test chambers used in the test similar in size, shape, and construction material? . . . . . 1__
4 Were the treatments randomly assigned to the test chamber locations? . . . . . 3__
5 Were the test organisms either randomly or impartially distributed to the test chambers? . . . . . 3__

C. Dilution Water

1 Did a daphnid, ceriodaphnid, mysid, or the test species survive for 30 or more days in the dilution water in the laboratory? . . . . . 3__
2 Was the dilution water obtained from the same source and pretreated in the same way throughout the test? . . . . . __

D. Test Material

1 Was the test material reagent-grade or better? . . . . . __
2 Is it true that undissolved test material did not cause a film, sheen, globules, etc., on top of the test solutions at the beginning of the test? . . . . . 40__
3 If the test material contained a cation of a heavy metal and if one or more of the tested concentrations were below the solubility limit of the material, is it true that none of these test solutions contained undissolved test material? . . __
4 Were the stock solution and the test solutions prepared without using a surfactant? . . . . . __
5 Were the stock solution and the test solutions prepared without using a solvent other than water? . . . . . 1__

Enter a dash for items D6 and D7 if the entry for item D5 was a dash.

6 If a water-miscible solvent was used, was it reagent-grade or better? . . . . . __
7 If a water-miscible solvent was used, was its concentration ≤ 0.5 mL/L (0.05%) in each important treatment? . . . . __

E. Test Conditions

1 Was it a flow-through test? . . . . . __
2 Was the test solution in each of the important test chambers sufficiently calm that the test organisms were not moved by current and/or turbulence? . . . . . 1__
3 Was the range in measured concentrations of test material ≤ a factor of 1.5 in each individual important test chamber? . . . . . __
4 Was the range in measured pH ≤ 0.5 in each individual important test chamber? . . . . . __
5 Was the range in all of the temperatures that were measured in important test chambers ≤ 2°C? . . . . . __
6 If the test was in fresh water, was the range in all of the measured hardnesses ≤ the higher of (a) 5 mg/L or (b) 10% of the average? . . . . . __
7 If the test was in salt water, was the range in all of the measured salinities ≤ the higher of (a) 2 g/kg or 20% of the average? . . . . . __
8 Were all of the concentrations of DO that were measured in important test chambers ≥ 6.5 mg/L? . . . . . __
9 Were all of the concentrations of dissolved nitrogen that were measured in important test chambers ≤ 103% of saturation? . . . . . __

F. Test Results

1 Enter 10U if it is not clear whether the results were based on concurrent effects, delayed effects, or both. . . . . __
2 If the test organisms were worms, is it true that reproduction did not occur in any test chambers? . . . . . __
3 Were the test organisms in each control treatment free of signs of disease, injury, and stress during the test? . . . . __
4 Was the % of test organisms that died or were affected in each control treatment ≤ 5%? . . . . . __
5 Was there a total of at least 40 test organisms in the important treatments at the beginning of the test? . . . . . __
6 If an LC50 or EC50 was calculated, was the % of test organisms that died or were affected in the lowest tested concentration that was not a control treatment ≤ 37%? . . . . . __
7 If an LC50 or EC50 was calculated, was the % of test organisms that died or were affected in the highest tested concentration ≥ 63%? . . . . . __
8 Can the relationship of EC50 (or LC50, etc.) to time be fit reasonably well with a smooth curve? . . . . . 3__

G. Measurement of Concentrations of Test Material in Test Solutions

Use Subsections G1, G2, and G3 below and the corresponding instructions to determine the entry for G.

Subsection G1: Aspects of the Measurement Process

G1a. Was the analytical method sufficiently selective? . . . . . __
G1b. Was the concentration of test material determined in appropriate samples of test solutions? . . . . . __
G1c. Were samples of test solutions handled (obtained, stored, shipped, etc.) acceptably? . . . . . __
G1d. Were at least 10% of the samples spiked and analyzed and was the mean recovery ≥ 87% and ≤ 115%? . . . . . __
G1e. Were at least 10% of the samples analyzed twice and was the geometric mean of the quotient of the high divided by the low ≤ 1.15? . . . . . __
G1f. Was a reference standard used? . . . . . __

G1 = G1a + G1b + G1c + G1d + G1e + G1f = __

Subsection G2: Special Conditions

G2a. Volatility

1. Enter the estimated number of hours it takes for 50% of the test material to volatilize from still water at 25°C. If the estimate is ≥1000 hr, enter A=1000 and enter G2a=0; then go to Item G2b. . . . . A = __
2. Enter the average residence time (in hours) of the test solutions in the test chambers. . . . . B = __
3. Enter the mean test temperature (in °C). . . . . C = __
4. Enter 2 if the test solutions were aerated and 1 if they were not. . . . . D = __
5. Enter 1000 if the test solutions were prepared by evaporating stock solution onto a surface of the test chambers and adding dilution water; otherwise enter 0. . . . . E = __

$$G2a = [20\{1-e^{-(\ln 2)(B/A)}\}^{0.6}][(1.414)^{\{(C-25)/5\}}][D]+[E/(4A)] = \_\_$$

G2b. Hydrolysis

1. Enter the estimated number of hours it takes for 50% of the test material to hydrolyze at 25°C. If the estimate is ≥1000 hr, enter F=1000 and enter G2b=0; then go to Item G2c. . . . . F = __
2. Enter the average residence time (in hours) of the test solutions in the test chambers. . . . . G = __
3. Enter the mean test temperature (in °C). . . . . H = __

$$G2b = [20\{1-e^{-(\ln 2)(G/F)}\}^{0.6}][(1.1)^{(H-25)}] = \_\_$$

G2c. Biodegradation

1. Enter the estimated number of hours it takes for 50% of the test material to biodegrade at 25°C. If the estimate is ≥1000 hr, enter J=1000 and enter G2c=0; then go to Item G2d. . . . . J = __
2. Enter the average residence time (in hours) of the test solutions in the test chambers. . . . . K = __
3. Enter 2 if the test solutions were aerated and 1 if they were not. . . . . L = __
4. Enter 2 if the test chambers contained a substrate that contained organic carbon and 1 if they did not. . . . . M = __

$$G2c = [20\{1-e^{-(\ln 2)(K/J)}\}^{0.6}][L][M] = \_\_$$

G2d. Sorption of organic chemicals

1. If the test material is an inorganic chemical, enter N=0. If the test material is an organic or organometallic chemical, enter N=log Kow. If N<4, enter G2d=0; then go to Item G2e. . . . . N = __
2. Enter the average residence time (in hours) of the test solutions in the test chambers. . . . . P = __
3. Enter 1 if the test chambers contained a substrate that contained organic carbon and 0 if they did not. . . . . Q = __
4. Enter 1 if the test organisms were fed during the test and 0 if they were not. . . . . R = __
5. Enter 1 if any of the surfaces (e.g., test chambers or liners) in contact with the test solutions during the test were a material (e.g., PVC, wood, concrete, etc.) that sorbs many test materials; enter 0 if all surfaces in contact with test solutions were glass, stainless steel, or a fluorocarbon plastic. . . . . S = __

$$G2d = (N-4)^3(P)(1+Q+R+S) = \_\_$$

G2e. Conditions that can affect the toxicities of cations of heavy metals

1. Enter 1 if the test material was a cation of a heavy metal; otherwise, enter T=0 and enter G2e=0; then go to Subsection G3. . . . . T = __
2. Enter the average resident time (in hours) of the test solutions in the test chambers. . . . . U = __
3. Enter 4 if the test chambers contained a substrate that contained organic carbon and 0 if they did not. . . . . V = __
4. Enter 4 if the test organisms were fed during the test and 0 if they were not. . . . . W = __
5. Enter 4 if any of the surfaces (e.g., test chambers or liners) in contact with the test solutions during the test were a material (e.g., PVC, wood, concrete, etc.) that sorbs many test materials; enter 0 if all such surfaces were glass, stainless steel, or a fluorocarbon plastic. . . . . X = __
6. Enter 4 if the test solutions were prepared by evaporating stock solution onto a surface of the test chambers and adding dilution water; otherwise enter 0. . . . . Y = __

$G2e = (T)(U)^{0.5}(V+W+X+Y) =$ __

G2 = 1+(G2a)+(G2b)+(G2c)+(G2d)+(G2e) = __

Subsection G3: Was the test result (or concentration-time-effect data) in EVISTRA based on measured concentrations of the test material? . . . . . G3 = __

Entry for G (maximum weight is 40) . . . . . = [(G1)(G2)]+[G3] = __

H. Miscellaneous . . . . . __

Total Number of Ns _____ **Total Number of Us** _____ **Total Number of Ms**_____

Comments_____

INSTRUCTIONS FOR USING THE ACUTE QUALITY CHECKLIST

A set of general instructions is given, and then specific instructions are given concerning individual items on the checklist.

General Instructions

1. This checklist should be applied to acute toxicity test results individually unless two or more acute test results in the same document are so similar that they can be addressed together because all of the entries are the same for the two or more test results. (It is very unlikely that two test results will have identical entries for all items on the checklist.)
2. There must be an entry for each item on the checklist; do not leave any blanks.
3. An N should be based on information that is known about a test result; an N should not be based on a lack of information.
4. Enter U if insufficient relevant information is available concerning the item. When EPA wants tests evaluated for EVISTRA, assumptions should not be made unless approved by the EVISTRA Coordinator. For example, if a document gives results of both acute and chronic tests and gives much information concerning the method used for the chronic test, but gives very little information concerning the method used for the acute test, it should not be assumed that information concerning the acute test can be extrapolated from the chronic test.
5. There are two situations in which the entry cannot be M:
   a. M must not be used to refer to information in another document by the same author(s); such information is considered "primary information" because it is by the same author(s). The other document by the same author(s) must be read to verify what it says and the other document must be cited as a comment at the end of the checklist even if it is cited in the document being reviewed. When the information is in another document by the same author(s), enter a dash, N, or U as appropriate; do not enter M.
   b. Some items, such as those concerning "Test Conditions" and "Test Results", refer to information that is produced by the test and therefore the entry cannot be M.
6. When a weight is calculated using an equation, restrictions are usually placed on use of the equation at both the low and high ends. In addition, the maximum weight allowed for any item is 40. Also, if the calculated weight is rounded, sufficient digits must be carried to prevent rounding from affecting the summary scores. Therefore, at least two decimal digits (not necessarily two significant digits) must be retained in all calculated weights to prevent rounding (including the possible use of different rules for rounding) from affecting the summary score. (Apparently some computers simply drop digits, instead of applying a rule for rounding.)
7. Some items require use of a range or standard deviation (SD). In some cases, the range or SD will be reported or can be calculated from the raw data. In other

cases, it will be necessary to convert between the range, SD, standard error (SE) of the mean, and/or relative standard deviation (RSD) of a measured water quality characteristic. If conversion is necessary, the following instructions should be followed.

a. The RSD is the quotient of the standard deviation (SD) divided by the mean; thus the RSD is the same as the coefficient of variation (COV).

b. If it is necessary to estimate a SD from a range, the SD should be estimated by multiplying the range by the factor given in the following table (Snedecor and Cochran 1967), where n is the number of measurements in the dataset:

n	factor	n	factor	n	factor
2	0.886	8	0.351	18	0.275
3	0.591	9	0.337	20	0.268
4	0.486	10	0.325	30	0.245
5	0.430	12	0.307	40	0.231
6	0.395	14	0.294	50	0.222
7	0.370	16	0.283	99	0.200

If n is unknown and there is no good basis for estimating n, enter U.

The factor depends on n because the range, but not the SD, is likely to increase as the number of measurements increases. [A brief version of this table is given on page 61 of Sokal and Rohlf (1969).]

c. If the standard error (SE) of the mean is reported, the SD should be calculated using the formula: $SD = (SE)(n^{½})$, where n is the number of measurements.

d. If the mean is reported and the range is reported to be, for example, a factor of 2, the range should be assumed to be centered on the mean. Similarly, if the range is reported to be from A to B, the mean should be assumed to be (A+B)/2.

e. If the results of measurements are reported in the x±y format:
   1. If it is known that the upper and lower limits of the range are x+y and x-y, the range is 2y, from which the SD can be estimated.
   2. If it is known that y is the SD, the SD does not have to be calculated or estimated.
   3. If it is known that y is the SE of the mean, the SD should be calculated using the equation given above.

   If the results are reported in the x±y format but the definition of y is not clear, either the appropriate number of Us should be assigned or the worst-case assumption should be made that the range is 2y, depending on which option results in the lowest negativity score.

f. Because the highest measured value tends to increase as the number of measurements increases and the lowest measured value tends to decrease as the number of measurements increases, use of the highest and/or lowest measured value in the evaluation of test results might discourage some people

from making more than the minimum number of required measurements. To avoid the possibility of discouraging people from making additional measurements, weights are not based on the highest and/or lowest measured values; rather, weights are based on values that are calculated using 1.96, which is the two-tailed value of Z at a probability of 0.95 (Snedecor and Cochran 1980; Sokal and Rohlf 1969). Thus the higher value used is m+1.96SD and the lower value used in m-1.96SD, where m is the mean and SD is the standard deviation.

8. In a toxicity test with aquatic organisms, test chambers are defined as the smallest physical units between which there are no water connections because treatments can be independently assigned to physical units between which there are no water connections, but they cannot be independently assigned to units between which there are water connections.

   Two endpoints must not be stored in EVISTRA from duplicate sets of chambers in the same acute toxicity test. If two endpoints were reported from duplicate sets of chambers in the same acute test, do one of the following:
   a. If raw data are available, they should be entered into EVISTRA.
   b. If raw data are available but are not entered into EVISTRA, the raw data must be used to calculate a new endpoint that is then entered into EVISTRA.
   c. If raw data are not available, the geometric mean of the two endpoints must be entered into EVISTRA as the endpoint.
9. The “important treatments” are the following:
   a. If the LC50 or EC50 is a “greater than” value, the important treatments are the highest tested concentration and all of the tested concentrations that were within a factor of two of the highest tested concentration.
   b. If the LC50 or EC50 is a “less than” value, the important treatments are the lowest tested concentration and all of the tested concentrations that were within a factor of two of the lowest tested concentration.
   c. If a point estimate of the LC50 or EC50 is calculated, the important treatments are all of the tested concentrations that were within a factor of two above or below the LC50 or EC50.

   These are the treatments that provide the critical information about the LC50 or EC50. Other treatments can provide confirmatory information, but such information is not as useful as that provided by the “important treatments”.

   Each test chamber that contains an “important treatment” is an “important test chamber” as long as it contains live test organisms, if the chamber is used in the calculation of the results for a treatment or test.
10. The “matrix ions” in both fresh and salt water are calcium, magnesium, sodium, potassium, chloride, sulfate, carbonate, and bicarbonate.
11. One of three techniques is used when an acute test is conducted:
    a. In the static technique, test solutions and organisms are placed in chambers and remain there for the duration of the test.

b. In the renewal technique, test organisms are exposed to “new” test solution of the same composition one or more times after the beginning of the test, usually once every 24 hours, either by transferring the organisms from one test chamber to another or by removing and replacing nearly all of the test solution in each test chamber.
c. In the flow-through technique, fresh test solution is added to the chamber continuously or every few minutes and excess solution flows out of the test chamber.

These are slightly modified from section 3.2 of ASTM Standard E729 (ASTM 1998b).

12. The following abbreviations are used:
    DO = dissolved oxygen
    DOC = dissolved organic carbon
    EDTA = ethylenediaminetetraacetic acid
    NTA = nitrilotriacetic acid
    POC = particulate organic carbon
    SD = standard deviation
    SE = standard error (of the mean)
    TCDD = tetrachlorordibenzo-*p*-dioxin
    TOC = total organic carbon
    TSS = total suspended solids
13. If there are several ways of determining an entry for an item, the instructions usually address the simplest ways first so that reviewers encounter the more complicated ways only when necessary.

Specific Instructions

A. Test Organisms

Item A1

a. Enter a dash if all of the test organisms hatched or were born at about the same time in the same hatchery or culture unit regardless of whether all of the reproducing adults in the hatchery or culture unit came from the same source because all of the organisms obtained from one hatchery or culture unit at one time are likely to be similar to each other.
b. Enter a dash if all of the organisms were from the same body of water (pond, reservoir, river, etc.), but enter N if they were from different bodies of water, even if the different bodies of water were in the same geographic area.
c. Enter U if all of the test organisms were from the same supply house, except enter a dash or N as appropriate if the supply house furnished information regarding its source(s) of the organisms.

Item A2

a. Enter a dash if the organisms were obtained from a supply house, hatchery, or culture unit.

b. Enter a dash if the organisms were obtained ten or more days prior to the start of the test, regardless of how they were obtained.
c. Enter U if, fewer than ten days prior to the start of the test, the organisms were collected from the field, but the method of collection is not known.

Item A3

If the range in age was a week or more and if the test species was a short-lived species or if the test organisms were young of a long-lived species, it is quite possible that the test organisms were in two or more life stages and the EVISTRA coordinator should be contacted for advice if the test results are being evaluated for EVISTRA.

For juvenile and adult fish and other organisms in a long-duration life stage (i.e., a life stage that lasts for more than five days):

a. Enter a dash if all of the test organisms were (1) in the same life stage at the start of the test and (2) the range of the weights of the organisms was not greater than a factor of two. (For fish, if the range of the cubes of the lengths is available but the range of the weights is not available, the range of the cubes of the lengths should not be greater than a factor of two.)
b. Enter N if some, but not all, of the organisms were in a long-duration life stage or if the organisms were all in the same long-duration life stage but the range of the weights of the organisms was more than a factor of two. (For fish, if the range of the cubes of the lengths is available but the range of the weights is not available, the range of the cubes of the lengths should not be greater than a factor of two.)
c. Enter U if no relevant information is available; for example, neither mean weight nor mean length provide information concerning the range.

For all other organisms:

A. Enter a dash if either (1) all of the test organisms were in the same short-duration life stage (i.e., a life stage that does not last for more than five days) at the start of the test, or (2) two successive life stages were used and it is quite difficult to separate them. For example, enter a dash if a test was conducted with two successive instars of an aquatic insect.
B. Enter N if the organisms were in three or more life stages.

Item A4

If the culture was examined for ephippial females at least once a week during the four weeks immediately prior to the start of the test:

a. Enter a dash if no ephippial females were found.
b. Enter N if one or more ephippial females were found.

Enter U if no information is available concerning the presence of ephippial females or if there is an explicit reason to believe that the culture was not examined for ephippial females at least once a week during the four weeks immediately prior to the start of the test.

Item A5

Enter a dash if (a) the test organisms were from a culture unit in the same facility, (b) the test organisms lived in the laboratory for at least the last ten days prior to the start of the test, or (c) the number of days was fewer than ten, but a side-by-side comparison using the same test material and the same species from the same water showed that that number of days gave the same result as ten or more days.

Item A6

Enter N if the organisms were subjected to either a change in temperature of more than 3°C in a 24-hr period and/or a change of more than 20 percent in hardness, alkalinity, conductivity, and/or salinity or more than 0.4 pH unit in a 24-hour period during the last ten days prior to the start of the test (or since birth or hatch if the organisms were less than ten days old).

Enter a dash if the only relevant information available is that the test organisms were acclimated.

Item A7

Enter a dash if it was reported that the test organisms were acclimated to the test chambers for at least the last 48 hours prior to the start of the test because it is reasonable to assume that the test chambers contained the dilution water, unless information to the contrary is available.

Enter 10U or 10M, as appropriate, if (a) the test organisms were cultured in the facility but it is not known whether they were cultured and tested in the same water or (b) the test organisms were obtained less than 48 hours prior to the start of the test and no information is available concerning the water that they were in prior to being obtained.

Otherwise, enter 10N, except (a) enter 3N if hardness, alkalinity, pH, and conductivity were similar for two fresh waters or (b) enter 3N if pH and salinity were similar for two salt waters. Two hardnesses, alkalinities, conductivities, or salinities are considered similar if the highest divided by the lowest is $\leq$1.1; two pHs are considered similar if the difference is $\leq$0.2.

Item A8

Enter a dash if the range of the temperature during acclimation was $\leq$3°C and the test temperature was within that range.

Enter N if the mean temperature during the test was not within 1°C of the mean temperature (a) for the last 48 hours prior to the start of the test or (b) since birth or hatch.

Enter a dash if it is known that the test organisms were acclimated unless there is a reason to enter N.

Item A9

Organisms that were acclimated acceptably were not necessarily handled properly and/or not subjected to substantial disturbance prior to and during the test.

Substantial disturbance can be either sustained disturbance or severe disturbance.

Enter a dash if, during the test and during the last 48 hours prior to the start of the test, the organisms were in culture chambers, holding chambers, test chambers, a test system, and/or a diluter system, unless there is an explicit reason to believe that the system was not protected from disturbance.

Enter N if there is an explicit reason to believe that the organisms probably were handled improperly and/or subjected to disturbance. (The EVISTRA coordinator will provide additional information as necessary concerning procedures used to generate test results that EPA wants evaluated for EVISTRA.)

If none of the above apply, enter U or M as appropriate.

Item A10

Enter 40U or 40M, as appropriate, if it is not known whether the test was static, renewal, or flow-through.

Loading limits are used to try to avoid stressing test organisms due to aggression, low concentrations of DO, and high concentrations of ammonia, but aggression and DO are addressed elsewhere. Therefore, regardless of the loading, enter a dash if the concentration of ammonia was measured at least daily and was below acceptable upper limits (WHICH HAVE NOT BEEN DEFINED YET).

If the loading is reported, the entry is assigned as follows for static tests whose duration was 96 hours, where L = the loading (in g/Liter of solution in the test chamber) and x = the number of Ns:

≤12°C for coldwater species;	>12°C for coldwater species;
≤17°C for warmwater species	>17°C for warmwater species
If L ≤ 0.8 g/L, enter a dash.	If L ≤ 0.5 g/L, enter a dash.
If L > 0.8 g/L, x = 12.5(L-0.8).	If L > 0.5 g/L, x = 20(L-0.5).

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:	For example:
If L = 0.88 g/L, enter 1.00N.	If L = 0.55 g/L, enter 1.00N.
If L = 1.6 g/L, enter 10.00N.	If L = 1.0 g/L, enter 10.00N.
If L = 3.2 g/L, enter 30.00N.	If L = 2.0 g/L, enter 30.00N.

For static tests whose durations were longer or shorter than 96 hours, 0.8 and 0.5 g/L are adjusted by multiplying by 96 and then dividing by the number of hours that the test organisms were in a batch of test solution, i.e., dividing by the duration of the test. For a 24-hr static test with a salmonid at 10°C, for example, 0.8 would be multiplied by (96/24) to give 3.2. Therefore:

If L $\leq$ 3.2 g/L, enter a dash.
If L > 3.2 g/L, x = 12.5(L-3.2).

For flow-through tests, the volume of solution in the test chamber is the average volume over a period of time such as 24 hours; unless the volume of solution in the test chamber fluctuates substantially, this is the volume of solution in the test chamber at any point in time. The values of 0.8 and 0.5 g/L are adjusted by multiplying by the flow rate, expressed in terms of the number of volume additions per day. If the flow rate is reported in terms of the time to achieve a specified percent replacement of the test solution, the corresponding number of volume additions per day should be determined using Figure 1 in Sprague (1973). For example, if the time to achieve 95 percent replacement was 3 hours, the number of volume additions per hour would be 1 and the number of volume additions per day would be 24. If the test was with a salmonid at 10°C, 0.8 would be multiplied by 24 to give 19.2. Therefore:

If L $\leq$ 19.2 g/L, enter a dash.
If L > 19.2 g/L, x = 12.5(L-19.2).

For renewal tests:

a. If part of the test solution was removed and replaced, use the percentage of the solution that was renewed and the average number of hours between renewals to determine the corresponding number of volume additions per day using Figure 1 in Sprague (1973).
b. If the test organisms were netted and transferred to fresh test solution, treat it as a static test whose duration equaled the average time between transfers.

Enter 40U if insufficient relevant information (e.g., loading, weight of the organisms, volume of test solution, and/or concentration of ammonia) is available. However:

a. It might be possible to estimate weight if length and/or age is reported. Estimated weights must not be low and must not be excessively high. Approval must be obtained from the EVISTRA coordinator before estimated weights are used when EPA wants test results evaluated for EVISTRA.
b. If it is known that the acute test was conducted in the same diluter system as a chronic test and there is no mention of modification of the system between the tests, it seems reasonable to assume that the flow rate was the same for the two tests.

If M, enter 40M.

Item A11

Enter a dash only if there is an explicit reason to believe that no aggression occurred.

Enter N if there is an explicit reason to believe that aggression occurred.

Otherwise, enter U.

Do not enter M.

Item A12

Test organisms can be stressed, even to the point of death, if food is withheld for too long. The maximum number of consecutive days that it is acceptable to not give food to the test organisms depends on the species, life stage, and/or size of the organisms. Extreme lack of food can reduce survival in the control treatments, but organisms can be stressed before they die. Unfortunately, it is usually difficult to quantify the period without food because it is often unclear how long food was available to the organisms after they were fed. In addition, the number of days without feeding that causes stress might depend on the quality and quantity of feed provided when the organisms were fed.

If sufficient relevant information is available, enter a dash if the maximum number of days was no more than fifty percent of the number of days that it would take for the test organisms to start dying after the last feeding, but enter N if they had not been fed for a longer period of time. Thus, it is possible that the entry for a test result obtained near the beginning of a test would be a dash, whereas the entry for a test result obtained near the end of the same test would be N.

If there is an explicit reason to believe that the organisms were fed up to the beginning of the test (e.g., during acclimation), it seems reasonable to assume that they were fed within 24 hours of the beginning of the test, unless information to the contrary is available.

If it is not known how long food can be withheld before the organisms would start dying due to lack of food, the entry is assigned as follows:

a. For saltwater mysids, enter a dash if they were fed at least daily before and during the test, but enter N if they were not fed at least daily before and during the test.
b. For daphnids and larvae of midges and phantom midges, enter a dash if food was withheld for no more than two days, but enter N if food was withheld for more than two days.
c. For all other invertebrates whose mean wet weight (blotted dry) was $\leq$ 0.5 g, enter a dash if food was withheld for no more than four days, but enter N if food was withheld for more than four days.
d. For invertebrates whose mean wet weight (blotted dry) was > 0.5 g, enter a dash if food was withheld for no more than five days, but enter N if food was withheld for more than five days.
e. For amphibian larvae and fish that had been actively feeding for fewer than 20 days, enter a dash if food was withheld for no more than four days, but enter N if food was withheld for more than four days.
f. For amphibian larvae and fish that had been actively feeding for at least 20 days and whose mean wet weight (blotted dry) was $\leq$ 0.5 g, enter a dash if food was withheld for no more than five days, but enter N if food was withheld for more than five days.

g. For amphibian larvae and fish that had been actively feeding for at least 20 days and whose mean wet weight (blotted dry) was > 0.5 g, enter a dash if food was withheld for no more than six days, but enter N if food was withheld for more than six days.

Enter U if needed information (e.g., whether or how long food was withheld from the test organisms, the size or age of the test organisms, etc.) is not available.

Item A13

The exposure might have been planned, such as for the treatment of a disease, or unplanned, such as if the unusual substance occurred in a surface water from which the organisms were obtained or in a reconstituted water in which the organisms were held.

The following are not considered unusual substances:

a. Water-soluble substances that consist only of hydrogen ion, hydroxide ion, water, and matrix ions (i.e., calcium, magnesium, potassium, sodium, chloride, sulfate, bicarbonate, and carbonate), as long as the concentrations of the matrix ions in the dilution water were not more than a factor of 4 above the average concentrations in the ambient water in which the test organisms usually live in field situations.
b. A commercially available sea salt (or a defined mixture of chemicals that is formulated to simulate sea salt), when the test was conducted with a saltwater species.
c. A food for the test organisms.
d. Surfactants and solvents (because they are addressed elsewhere).
e. Substances that have been demonstrated in side-by-side tests to have no effect on the result of a toxicity test on the test material. The side-by-side tests may be from a different document but must have used (1) a species in the same phylum as the test species in the toxicity test being evaluated and (2) concentrations similar to those used in the toxicity test being evaluated.

As necessary, the EVISTRA coordinator will provide detailed guidance concerning each unusual substance related to test results that EPA wants evaluated for EVISTRA.

Except when they are covered by “a” and/or “e” above, the following are examples of substances that are considered unusual substances:

1. Chemicals used to treat test organisms for diseases or parasites.
2. Buffers.
3. High concentrations of TOC, TSS, and color, except that a high concentration of TOC is not an unusual substance if the concentration of TOC is acceptably measured and reported.
4. EDTA, NTA, citrate, and other complexing agents.
5. The test material, except that for such test materials as copper and zinc, background concentrations in the dilution water are not considered unusual substances.

6. Any other substance whose concentration in the dilution water is greater than its usual concentration in the ambient water in which the test organisms usually live in field situations. (As necessary, the EVISTRA coordinator will provide detailed guidance concerning such substances related to test results that EPA wants evaluated for EVISTRA.)

Enter 40N if there is an explicit reason to believe that the dilution water contained one or more of the unusual substances mentioned above in situations in which the test organisms were in the dilution water prior to the start of the test.

If the dilution water did not contain an unusual substance but the organisms were exposed to one or more unusual substances during the last ten days prior to the start of the test, the number of Ns is assigned as follows, where d = number of days in clean water immediately prior to the test and x = the number of Ns:

$x = (40/10)(10-d)$ for $0 \leq d < 10$

If $d \geq 10$, enter a dash.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:

If d = 0, enter 40.00N

If d = 7, enter 12.00N

Unless there is a reason to enter N, enter a dash if the test organisms were held for 10 or more days in the same facility or were from a culture unit in the same facility.

If none of the above apply, enter 40U or 40M as appropriate.

Item A14

See A13 concerning the definition of “unusual substance”.

Enter N if there is an explicit reason to believe that the dilution water contained “unusual substances” during the test.

Enter a dash if there is no explicit reason to enter N.

B. Experimental Design

Item B1

Even if there is an explicit reason to believe that the acute test was conducted in the same diluter system as a chronic test and there is no mention of modification of the system between the tests, it should not be assumed that the two tests had the same number of treatments because test organisms might not have been placed in all of the test chambers for the acute test.

Enter 3N if any of the following are true:

a. There was no control treatment.

b. The “control treatment” was not comparable to the other treatments. For example, organisms in a holding tank, culture unit, or pond are not an acceptable control treatment.
c. A solvent that was not the test material was used in the preparation of the stock solution and/or the test solutions, but (1) there was no solvent control treatment or (2) the concentration of solvent in the solvent control treatment was lower than the highest concentration of solvent in another treatment.
d. A surfactant (or another chemical that is not a solvent and was not the test material) was used in the preparation of the stock solution and/or the test solutions, but (1) no control treatment contained the surfactant or chemical or (2) the concentration of the surfactant or chemical was not the same in all of the treatments.

Enter 1N if there was a solvent, surfactant, or chemical control treatment, but no dilution-water control treatment.

If none of the above apply, enter a dash, 3M, or 3U as appropriate.

Item B2

If there is an explicit reason to believe that the acute test was conducted in the same diluter system as a chronic test and there is no mention of modification of the system between the tests, it seems reasonable to assume that the dilution factor was the same for the two tests. In contrast, it should not be assumed that the two tests had the same number of test chambers per treatment because test organisms might not have been placed in all of the test chambers for the acute test.

Item B3

Enter N if there is an explicit reason to believe that all of the test chambers were not identical. Enter a dash if it was reported that, for example, the test chambers were 500-mL beakers because it is reasonable to assume that all of the chambers were 500-mL beakers. Enter a dash if there is an explicit reason to believe that a diluter was used because it is reasonable to assume that all of the test chambers were identical.

If none of the above apply, enter U or M as appropriate.

Item B4

Information concerning other aspects of the test (e.g., it was a static test; a dilutor was used) provide no information concerning whether the treatments were randomly assigned to test chamber locations.

If there is an explicit reason to believe that the acute test was conducted in the same diluter system as a chronic test and there is no mention of modification of the system between the tests, it seems reasonable to assume that the assignment of treatments to test chamber locations was the same for the two tests.

Enter N if there is an explicit reason to believe that the treatments were not randomly assigned to the test chamber locations.

Enter a dash if there is an explicit reason to believe that, for example, either a complete randomized design or a randomized block design was used. (In a randomized block design, each block contains one test chamber for each treatment and the treatments are randomly assigned to test chamber locations within each block.)

Enter U if the only relevant information available is a vague statement concerning assignment of the treatments, e.g., "the treatments were arbitrarily assigned" or "the treatments were indiscriminately assigned."

If none of the above apply, enter U or M as appropriate.

Item B5

Enter a dash if the test organisms were assigned to the test chambers using one of the following procedures:

a. Test organisms were randomly assigned using random numbers or by drawing numbers from a "hat".
b. Twenty test organisms were impartially assigned by assigning a group of four organisms to each chamber, assigning another group of four organisms to each chamber, and repeating this process until each chamber contains 20 organisms.

Enter N if there is an explicit reason to believe that the test organisms were not assigned to the test chambers using one of the two procedures.

Enter U if the only relevant information available is a vague statement concerning assignment of the test organisms, e.g., "the organisms were arbitrarily assigned" or "the organisms were indiscriminantly assigned."

If none of the above apply, enter U or M as appropriate.

C. Dilution Water

Item C1

Enter a dash if one of the listed species was either (a) cultured or (b) held for 30 or more days in the dilution water in the laboratory during the time when the toxicity test was conducted. The relevant information may come from a document other than the report of the toxicity test being reviewed, but the other document must be cited as a comment at the end of the quality checklist.

If there is an explicit reason to believe that the test organisms were cultured in the same facility and that the acute test was conducted in a dilution water that might have been used in the culture of the test organisms, it seems reasonable to assume that the organisms were cultured and tested in the same water, unless information is available to the contrary.

Otherwise, enter N or U as appropriate.
Do not enter M.

Item C2

a. Enter a dash if either (a) it was a static test or (b) the dilution water used in the test was obtained from the same source and received either no treatment or the same treatment throughout the test.
b. Enter 20N if it was reported that the dilution water was obtained from two or more sources and/or received different treatments during the test, but there was an attempt to ensure that the dilution water had similar water quality characteristics throughout the test.
c. Enter 40N if it was reported that the dilution water was obtained from two or more sources and/or received different treatments during the test and it is likely that the water quality characteristics of the dilution water changed substantially during the test.
d. Do not enter U or M.

D. Test Material and Test Solutions

Item D1

Information concerning an analytical standard used in chemical analyses does not qualify as information concerning the test material.

The test material should be reagent-grade or better to reduce the possibility that an impurity might affect the results of the test. Enter a dash if:

a. The test material is reported to be reagent-grade, ACS-grade, analytical grade, or a reference standard.
b. Similar results have been obtained in side-by-side tests on test material that is not reagent-grade or better as on test material that is reagent-grade or better. The side-by-side tests may be from a different document but must have used a species in the same phylum as the test species in the toxicity test being evaluated. (Results of such comparisons should usually be extrapolated from one low-purity material to another only if it is likely that the impurities are similar.) As necessary, the EVISTRA coordinator will provide detailed guidance concerning each test material for which EPA wants results evaluated for EVISTRA.

If neither of the above apply and the purity of the test material is reported, the entry is assigned as follows, where P = the percent purity and x = the number of Ns:

If $P \geq 99$ %, enter a dash.
If $P < 99$ %, $x = 2(99 - P)$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:

If P = 98, enter 2.00N.
If P = 95, enter 8.00N.
If P = 89, enter 20.00N.

Enter 20N if there is an explicit reason to believe that the test material was technical grade and no other relevant information is available.

Enter 40N if the test material was an emulsifiable concentrate, wettable powder, or specially prepared mixture because of the possibility that an impurity might affect the results of the test, except enter a dash if similar results have been obtained in side-by-side tests on that test material as on test material that is reagent-grade or better. (Results of such comparisons should usually be extrapolated from one formulated material to another only if it is likely that the ingredients are similar.) As necessary, the EVISTRA coordinator will provide detailed guidance concerning each test material for which EPA wants results evaluated for EVISTRA.

Enter 40U if no information is available concerning the quality of the test material.

Item D2

a. Enter 40N if there is an explicit reason to believe that there was a film, sheen, globules, or some other form of undissolved test material on the surface of the test solutions at the beginning of the test through which the test organisms would have had to pass to be placed into the test chambers.
b. Otherwise enter a dash.

Item D3

a. If the test material contained a cation of a heavy metal and the concentration of test material was below the solubility limit of the test material, but the test solution contained undissolved test material, enter 40N because there is a very large problem. As necessary, the EVISTRA coordinator will provide detailed guidance concerning the solubility limit of each test material for which EPA wants results evaluated for EVISTRA.
b. Enter a dash if it was reported that no test chamber contained undissolved test material.
c. Otherwise enter U.
d. Do not enter M.

Item D4

It is important to correctly differentiate between surfactants and solvents because surfactants have a large effect on the surface tension of test solutions, whereas solvents do not. Carriers and additives are usually either surfactants or solvents.

An organic chemical is miscible with water if it is a liquid that can be mixed with water in all ratios without separation into two phases. An organic chemical is not miscible with water is there is any ratio at which a mixture separates into two phases. As necessary, the EVISTRA coordinator will provide detailed guidance concerning the miscibility (with water) of each liquid organic chemical for which EPA wants test results evaluated for EVISTRA.

If the test material was either an organic chemical that is miscible with water or an inorganic chemical:

1. Enter a dash if there is no indication that a surfactant was used.

2. Enter 20N if available information indicates that a surfactant was used but there is no indication that similar results have been obtained in side-by-side tests on that test material using the same or a similar surfactant as on the test material without the surfactant.
3. Enter a dash if available information indicates that a surfactant was used and that similar results have been obtained in side-by-side tests on that test material using a similar or higher concentration of the same or a similar surfactant as on that test material without using the surfactant. The side-by-side tests may be from a different document but must have used a species in the same phylum as the test species in the toxicity test being reviewed. (Surfactants are considered similar only if they are structurally similar and their minor ingredients are similar.)

If the test material was an organic chemical that is not miscible with water, enter a dash only if either (a) it is explicitly stated that no surfactant was used or (b) the preparation of stock and test solutions is described in detail and no surfactant is mentioned.

Otherwise, enter 20U or 20M as appropriate.

Item D5

An organic chemical is miscible with water if it is a liquid that can be mixed with water in all ratios without separation into two phases. An organic chemical is not miscible with water is there is any ratio at which a mixture separates into two phases. As necessary, the EVISTRA coordinator will provide detailed guidance concerning the miscibility of each liquid organic chemical that was used in toxicity tests which EPA wants evaluated for EVISTRA.

If the test material was either an organic chemical that is miscible with water or an inorganic chemical:

1. Enter a dash if there is no indication that a solvent was used.
2. Enter 1N if available information indicates that a solvent was used but there is no indication that similar results have been obtained in side-by-side tests on that test material using the same or a similar solvent as on the test material without the solvent.
3. Enter a dash if available information indicates that a solvent was used and that similar results have been obtained in side-by-side tests on that test material using a similar or higher concentration of the same or a similar solvent as on that test material without using the solvent. The side-by-side tests may be from a different document but must have used a species in the same phylum as the test species in the toxicity test being reviewed. (Solvents are considered similar only if they are structurally similar and their minor ingredients are similar.)

If the test material was an organic chemical that is not miscible with water, enter a dash only if either (a) it is explicitly stated that no solvent was used or (b) the preparation of stock and test solutions is described in detail and no solvent is mentioned; otherwise, enter 1U or 1M as appropriate.

As necessary, the EVISTRA coordinator will provide detailed guidance concerning each solvent used in tests that generated results that EPA wants evaluated for EVISTRA.

Item D6

Enter a dash if the entry for D5 is a dash.
Enter 10U if the entry for D5 is U.

A solvent is miscible with water if it can be mixed with water in all ratios without separation into two phases. A solvent is not miscible with water is there is any ratio at which a mixture separates into two phases. Acetone, acetonitrile, dimethylformamide (DMF), ethanol, methanol, and triethylene glycol (TEG) are water-miscible organic solvents that are commonly used to prepare test solutions that contain organic chemicals.

The solvent should be reagent-grade or better to reduce the possibility that an impurity might affect the results of the test. Enter a dash if:

a. The solvent is reported to be reagent-grade, ACS-grade, analytical grade, or a reference standard.
b. Similar results have been obtained in side-by-side tests on that test material using a similar or higher concentration of a batch of the solvent that is not reagent-grade or better as on that test material using the same concentration of a batch of the solvent that is reagent-grade or better. The side-by-side tests may be from a different document but must have used a species in the same phylum as the test species in the toxicity test being reviewed. (Results of such comparisons should be extrapolated from one batch of a solvent to another batch of the same solvent only if it is likely that the impurities are similar.) As necessary, the EVISTRA coordinator will provide detailed guidance concerning specific solvents used in tests that generated results that EPA wants evaluated for EVISTRA.

If neither of the above apply and the purity of the solvent is reported, the entry is assigned as follows, where P = the percent purity and x = the number of Ns:

If $P \geq 99$ %, enter a dash.
If $P < 99$ %, $x = 2(99-P)$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:

If P = 98 %, enter 2.00N.
If P = 95 %, enter 8.00N.
If P = 90 %, enter 18.00N.

Enter 20N if there is an explicit reason to believe that the solvent was not better than technical grade and no other relevant information is available.

If M, enter 10M.
If unknown, enter 10U.

Item D7

Enter a dash if the entry for D5 is a dash.
Enter 10U if the entry for D5 is U.

The entry is assigned as follows, where C = the concentration of solvent in the important treatment containing the highest concentration of solvent and x = the number of Ns:

If $C \leq 0.5$ mL/L, enter a dash.
If $C > 0.5$ mL/L, $x = 5(C-0.5)$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:

If C = 0.6 mL/L, enter 0.50N.
If C = 1.0 mL/L, enter 2.50N.
If C = 2.5 mL/L, enter 10.00N.

If M, enter 10M.
If unknown, enter 10U.

E. Test Conditions

Item E1

Enter a dash if it was a flow-through test.
Enter 1N if it was a renewal test.
Enter 2N if it was a static test.
(The technique used will also affect other items.)

Item E2

Enter N if there is an explicit reason to believe that the test organisms were moved by a current and/or turbulence. If there is not an explicit reason to believe that the test organisms were moved by a current and/or turbulence, enter a dash if:

a. The test was static or renewal and the test solutions were not reported to have been aerated in the test chambers.
b. The test was an aerated static or renewal test and consideration of the rate and/or method of aeration, the volume of test solution, and the size of the test organisms implies that there would not have been sufficient current and/or turbulence to move the test organisms.
c. The test was a flow-through test with small swimming organisms (such as daphnids, ceriodaphnids, rotifers, and swim-up fish) and it was reported that a technique was used to keep turbulence low. (If it is reported that a technique was used to keep turbulence low, it should be assumed that turbulence was low, unless it was reported either that turbulence was high or that test organisms

were moved by current and/or turbulence.) A low flow rate is not sufficient by itself.

d. The test was a flow-through test with other organisms and the number of volume additions per day was less than 40.

A "volume addition" is the introduction into a test chamber of a volume of test solution equal to the average volume of solution in the chamber. If the flow rate is reported in terms of the time to achieve a specified percent replacement of the test solution, the corresponding number of volume additions per day should be determined using Figure 1 in Sprague (1973).

If it is known that the acute test was conducted in the same diluter system as a chronic test and there is no mention of modification of the system between the tests, it seems reasonable to assume that the flow rate and the average volume of test solution in the test chambers were the same for the two tests.

Enter U if none of the above apply; e.g., enter U if the number of volume additions per day was $\geq$40, but turbulence is not mentioned.

Do not enter M.

Item E3

It seems reasonable that, in a high quality test, the measured concentration of test material should not vary from one time to another in each individual important test chamber by more than a factor of 1.5.

Enter 40U if the most relevant information available is a statement that the measured concentration of test material was similar to the nominal (i.e., intended) concentration in a stock solution or in test solutions before they entered test chambers.

Enter 10U if the most relevant information available is a statement that the measured concentration of test material in each important test chamber was uniform during the test.

Enter 20N if the most relevant information available is a statement that the measured concentration of test material in one or more important test chambers varied substantially during the test.

If the instructions given above did not result in an entry, enter 40U if quantitative information is not available concerning the results of measurements of the concentration of test material in the test solutions in at least half of the important test chambers at two or more times, including near the beginning and end of the test. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. If the range in measured concentrations of test material in each individual important test chamber in which the concentration of test material was measured more than once was ≤ a factor of 1.5, enter a dash.

2. If an experimentally determined quotient is calculated for a test chamber by dividing the highest measured concentration of test material by the lowest measured concentration, the chances of exceeding 1.5 increase as the number of measurements increases. To make the chances of exceedence independent of the number of measurements, the quotient used here for a test chamber is defined as: Q = (m+1.96SD) / (m-1.96SD), where m = the arithmetic mean of all of the concentrations of test material that were measured in the test chamber and the standard deviation (SD) is derived from the same measured concentrations using the guidance presented in General Instruction #7. Variation in measured concentrations of test material is expected to be similar in important and unimportant test chambers and so variation in all test chambers should be used together. Nevertheless, the decision rules for rejecting outliers should make it easier to reject data from an unimportant test chamber than to reject data from an important test chamber.

   If sufficient information is available that Qs can be derived for two or more individual test chambers in which the concentration of test material was measured more than once, the entry is assigned as follows, where G = the geometric mean of all of the Qs for individual test chambers and x = the number of Ns:

   If G ≤ 1.5, enter a dash.
   If G > 1.5, x = $10^{(G-1.5)}$

   The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

   For example:
   If G = 1.6, enter 1.26N.
   If G = 2.0, enter 3.16N.
   If G = 3.0, enter 31.62N.

   If Qs cannot be derived for two or more individual test chambers in which the concentration of test material was measured more than once, but an arithmetic mean of chamber-specific or treatment-specific relative standard deviations (RSDs) is available for two or more test chambers in which the concentration of test material was measured more than once, the mean RSD should be multiplied by the acute value (e.g., LC50, EC50, etc.) and the product should be used as an estimate of the SD. The acute value should be used as an estimate of m and then H = (m+1.96SD) / (m-1.96SD) should be used as an estimate of G to determine the number of Ns, as described above.

Do not enter M.

Item E4

It seems reasonable that, in a high quality test, the measured pH should not vary from one time to another in each individual important test chamber by more than 0.5 pH unit.

Enter 20U if the most relevant information available concerns the dilution water or test solutions before they entered test chambers.

Enter 5U if the most relevant information available is a statement that the measured pH in each important test chamber was uniform during test.

Enter 10N if the most relevant information available is a statement that the measured pH in one or more important test chambers varied substantially during the test.

If the instructions given above did not result in an entry, enter 20U if quantitative information is not available concerning the results of measurements of the pH of the test solutions in at least half of the important test chambers at two or more times, including near the beginning and end of the test. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. If the range in measured pH in each individual important test chamber in which pH was measured more than once was $\leq$0.5 pH unit, enter a dash.

2. If an experimentally determined difference is calculated for a test chamber by subtracting the lowest measured pH from the highest measured pH, the chances of exceeding 0.5 increase as the number of measurements increases. To make the chances of exceedence independent of the number of measurements, the difference used here for a test chamber is defined as:

$$D = (m+1.96SD) - (m-1.96SD) = 3.92SD$$

where m = the mean of all of the values of pH that were measured in the test chamber and the standard deviation (SD) is derived from the same measured values of pH using the guidance presented in General Instruction #7. Variation in measured pH is expected to be similar in important and unimportant test chambers and so variation in all test chambers should be used together. Nevertheless, the decision rules for rejecting outliers should make it easier to reject data from an unimportant test chamber than to reject data from an important test chamber.

NOTE: Whenever a mean and SD (or standard error, SE) are reported for pH, it is important to determine how they were calculated. Sometimes they

are calculated from the measured values of pH, but sometimes they are calculated from the negative logs of the pH values and then transformed back to the pH units. (For the purpose of evaluating quality of a test result, it is not important how they were calculated as long as the procedure was used correctly for all of the values [e.g., mean, SD, SE, 95% confidence limits] that were calculated.) Although back transformation of a mean is appropriate, it is not appropriate to back transform a SD or SE; the correct procedure is to back transform upper and lower 95% confidence limits, even though the back-transformed limits will not be symmetrical around the back-transformed mean. Unfortunately, transformations are sometimes used incorrectly in the calculation of means, SDs, and SEs, and in back transformations. Therefore, even if the calculation of a mean and SD (or SE) is described carefully and appropriately, the mean and 95% confidence limits should be evaluated (e.g., compared with the original raw data) to determine if they are obviously incorrect. If they are obviously incorrect, they should not be used.

If sufficient information is available that Ds can be derived for two or more individual test chambers in which pH was measured more than once, the entry is assigned as follows, where A = the arithmetic mean of all of the Ds and x = the number of Ns:

If A $\leq$ 0.5 pH unit, enter a dash.
If A > 0.5 pH unit, x = 30(A-0.5).

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:
If A = 0.6, enter 3.00N.
If A = 1.0, enter 15.00N.
If A = 1.5, enter 30.00N.

If Ds cannot be derived for two or more individual test chambers in which pH was measured more than once, but B = (m+1.96SD) - (m-1.96SD) = 3.92SD can be derived (see General Instruction #7) from pooled data for two or more test chambers in which pH was measured more than once, B should be divided by 2 and the quotient should be used as an estimate of A, which is then used to determine the number of Ns, as described above.

Do not enter M.

Item E5

It seems reasonable that, in a high quality test, the measured temperature should not vary from one time in one important test chamber to another time in the same or a different important test chamber by more than 2°C.

Enter 20U if the most relevant information available either (a) is that the test was conducted at, for example, 20°C or (b) concerns the dilution water or test solutions before they entered test chambers.

Enter 5U if the most relevant information available is a statement that all of the temperatures that were measured in important test chambers during the test were similar.

Enter 10N if the most relevant information available is a statement that the measured temperature in one or more important test chambers varied substantially during the test.

If the instructions given above did not result in an entry, enter 20U if quantitative information is not available concerning the results of measurements of the temperature of the test solutions in at least half of the test chambers at two or more times, including near the beginning and end of the test. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. If the range in all of the temperatures that were measured in important test chambers was $\leq$2°C, enter a dash.

2. If an experimentally determined difference is calculated for a test by subtracting the lowest measured temperature from the highest measured temperature, the chances of exceeding 2°C increase as the number of measurements increases. To make the chances of exceedence independent of the number of measurements, the difference used here for a test is defined as:

   $$D = (m+1.96SD) - (m-1.96SD) = 3.92SD$$

   where m = the arithmetic mean of all of the temperatures that were measured at any time in any test chamber and the standard deviation (SD) is derived from the same measured temperatures using the guidance presented in General Instruction #7. Temperature is expected to be similar in important and unimportant test chambers and so all of the temperatures that were measured in all of the test chambers should be used together. Nevertheless, the decision rules for rejecting outliers should make it easier to reject data from an unimportant test chamber than to reject data from an important test chamber.

   If sufficient information is available that D can be derived from all of the measured temperatures, the entry is assigned as follows, where x = the number of Ns:

   If D $\leq$ 2°C, enter a dash.
   If D > 2°C, x = 5(D-2).

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:
If D = 2.1°C, enter 0.50N.
If D = 4°C, enter 10.00N.
If D = 6°C, enter 20.00N.

If D cannot be derived as described above because the only results available are chamber-specific summaries of measured temperatures, derive $F = (m+1.96SD) - (m-1.96SD) = 3.92SD$ for each test chamber and use the arithmetic mean of the Fs as an estimate of D, which is then used to determine the number of Ns, as described above.

Do not enter M.

Item E6

It seems reasonable that, in a high quality test, the measured hardness should not vary from one time in one important test chamber to another time in the same or a different important test chamber by more than the higher of (a) 5 mg/L or (b) 10% of the average measured hardness. In addition, it seems reasonable that the hardness (but not the pH, DO, and temperature) in the test chambers will be similar to that of the dilution water; therefore, measurement of the hardness of the dilution water can be used if hardness was not adequately measured in the test chambers.

Because this item applies only to freshwater tests in which new dilution water was used over time, enter a dash for all other tests; for example, enter a dash if the test was static and/or the dilution water was salt water.

Enter 5U if the most relevant information available is a statement that all of the hardnesses that were measured in important test chambers and/or all of the hardnesses that were measured in the dilution water during the test were similar.

Enter 10N if the most relevant information available is a statement that the hardness of the test solution in one or more important test chambers and/or the hardness of the dilution water varied substantially during the test.

If the instructions given above did not result in an entry, enter 20U if quantitative information is not available concerning the results of measurements of the hardness of the test solutions and/or the dilution water at two or more times, including near the beginning and end of the test. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. If the range in all of the measured hardnesses was ≤ the higher of (a) 5 mg/L or (b) 10% of the average measured hardness, enter a dash.

2. If an experimentally determined difference is calculated for a test by subtracting the lowest measured hardness from the highest measured hardness, the chances of exceeding the allowed range increase as the number of measurements increases. To make the chances of exceedence independent of the number of measurements, the difference used here for a test is defined as:

   $$D = (m+1.96SD) - (m-1.96SD) = 3.92SD$$

   where m = the arithmetic mean of all of the hardnesses that were measured and the standard deviation (SD) is derived from the same measured hardnesses using the guidance presented in General Instruction #7. Hardness is expected to be similar in important and unimportant test chambers and so all of the measured hardnesses should be used together. Nevertheless, the decision rules for rejecting outliers should make it easier to reject data from an unimportant test chamber than to reject data from an important test chamber.

   If sufficient information is available that D can be derived from all of the hardnesses that were measured in test chambers or all of the hardnesses that were measured in the dilution water, the entry is assigned as follows, where:
   T = the higher of (a) 5 mg/L or (b) 10% of the average measured hardness.
   x = the number of Ns.

   If D ≤ T, enter a dash.
   If D > T, x = 0.5(D-T).

   The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

   For example, for T=10 mg/L:
   If D = 12 mg/L, enter 1.00N.
   If D = 30 mg/L, enter 10.00N.
   If D = 70 mg/L, enter 30.00N.

   If D cannot be derived as described above because the only results available are chamber-specific summaries of hardness measurements, derive F = (m+1.96SD) - (m-1.96SD) = 3.92SD for each test chamber and use the arithmetic mean of the Fs as an estimate of D, which is then used to determine the number of Ns, as described above.

Do not enter M.

Item E7

It seems reasonable that, in a high quality test, the measured salinity should not vary from one time in one important test chamber to another time in the same or a different important test chamber by more than the higher of (a) 2 g/kg or (b) 20% of

the average measured salinity. In addition, it seems reasonable that the salinity (but not the pH, DO, and temperature) in the test chambers will be similar to that of the dilution water; therefore, measurement of the salinity of the dilution water can be used if salinity was not adequately measured in the test chambers.

Because this item applies only to saltwater tests in which new dilution water was used over time, enter a dash for all other tests; for example, enter a dash if the test was static and/or the dilution water was fresh water.

Enter 5U if the most relevant information available is a statement that all of the salinities that were measured in important test chambers and/or all of the salinities that were measured in the dilution water during the test were similar.

Enter 10N if the most relevant information available is a statement that the salinity of the test solutions in one or more important test chambers and/or the salinity of the dilution water varied substantially during the test.

If the guidance presented above did not result in an entry, enter 20U if quantitative information is not available concerning the results of measurement of the salinity of the test solutions and/or the dilution water at two or more times, including near the beginning and end of the test. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. If the range in all of the measured salinities was $\leq$ the higher of (a) 2 g/kg or (b) 20% of the average measured salinity, enter a dash.

2. If an experimentally determined difference is calculated for a test by subtracting the lowest measured salinity from the highest measured salinity, the chances of exceeding the allowed range increase as the number of measurements increases. To make the chances of exceedence independent of the number of measurements, the difference used here for a test is defined as:

   $$D = (m+1.96SD) - (m-1.96SD) = 3.92SD$$

   where m = the arithmetic mean of all of the salinities that were measured and the standard deviation (SD) is derived from the same measured salinities using the guidance presented in General Instruction #7. Salinity is expected to be similar in important and unimportant test chambers and so all of the measured salinities should be used together. Nevertheless, the decision rules for rejecting outliers should make it easier to reject data from an unimportant test chamber than to reject data from an important test chamber.

   If sufficient information is available that D can be derived from all of the salinities that were measured in test chambers or all of the salinities that were measured in the dilution water, the entry is assigned as follows, where:

T = the higher of (a) 2 g/Kg or (b) 20% of the average measured salinity.
x = the number of Ns.

If D $\leq$ T, enter a dash.
If D > T, x = 2(D-T).

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example, for T=6 g/Kg:
If D = 7 g/Kg, enter 2.00N.
If D = 11 g/Kg, enter 10.00N.
If D = 21 g/Kg, enter 30.00N.

If D cannot be derived as described above because the only results available are chamber-specific summaries of salinity measurements, derive F = (m+1.96SD) - (m-1.96SD) = 3.92SD for each test chamber and use the arithmetic mean of the Fs as an estimate of D, which is then used to determine the number of Ns, as described above.

Do not enter M.

Item E8

Unless the purpose of the test required a low concentration of dissolved oxygen (DO) in one or more test chambers, it seems reasonable that, in a high quality test, the concentration of DO in important test chambers should be sufficiently high that the test organisms are not stressed due to a low concentration of DO.

Enter a dash if the purpose of the test required a low concentration of DO in one or more test chambers.

Enter 10U if the most relevant information available is a statement that all of the concentrations of DO that were measured in important test chambers were near saturation.

A lower limit on the concentration of DO that is not expected to stress aquatic organisms can be expressed in a variety of ways:

1. A lower limit can be expressed in terms of either the percent of saturation (e.g., ASTM 1998b) or the concentration itself (e.g., Chapman 1986). Here, lower limits will be expressed in terms of the concentration, not the percent of saturation.
2. A lower limit can be expressed as a value that must be exceeded by each individual measured concentration of DO or as a value that must be exceeded by the mean of the measured concentrations. A value that must be exceeded by each measured concentration tends to penalize investigators who make more than the minimum number of measurements and ignores the fact that organisms

can tolerate very low concentrations of DO for brief periods of time. In contrast, a mean over the duration of the test can allow an excessively low concentration to exist for too long a period of time.

In order to provide adequate protection from stress without being unnecessarily restrictive, two kinds of lower limits will be used here:

a. A lower limit on m = the arithmetic mean of all of the DO concentrations measured in all of the test chambers containing live test organisms.
b. A lower limit on H = m-1.96SD, where SD is the standard deviation that is derived from the same set of measured DO concentrations using the guidance presented in General Instruction #7.

Both kinds of lower limits can be established for DO because much information is available concerning the effects of low concentrations of DO on aquatic organisms.

Because some types of test organisms are more sensitive to low concentrations of DO than others, the two lower limits will be specified for four types of test organisms:

Type of Organisms	Lower Limit on m (mg/L)	Lower Limit on H (mg/L)
Warmwater fishes		
Early life stages	6.0	5.0
Other life stages	5.5	4.0
Salmonids	6.5	5.0
Invertebrates	6.0	5.0

"Early life stages" include all embryonic and larval stages and all juvenile forms up to 30-days post hatch (U.S. EPA 1986). The first three lower limits on m are presented on page 34 of U.S. EPA (1986) and are 0.5 mg/L above the concentrations given for "slight production impairment" on page 31. U.S. EPA (1986) does not give a "mean" for invertebrates on page 34 and so the last lower limit on m given above is 1 mg/L higher than the concentration given for "some production impairment" of invertebrates on page 31. The lower limits on H are concentrations given on page 31 for "moderate production impairment" or "some production impairment".

Unless there was substantial aeration or substantial change in water temperature, the concentration of DO in test solutions in test chambers is expected to be equal to or lower than the concentration of DO in the dilution water or in the test solutions before they entered test chambers. Therefore, in the absence of substantial aeration or change in water temperature, a measured concentration of DO in dilution water and/or test solutions before they entered the test chambers can be used as an upper limit on the concentration of DO in the test chambers during the test.

If the most relevant information available concerns the concentration of DO in the dilution water and/or test solutions before they entered the test chambers:

1. Enter 20N if the concentration of DO in the dilution water and/or test solutions before they entered the test chambers was below the lower limit on H given above for the test organisms.
2. Otherwise, enter 30U.

The concentration of DO that occurs in test chambers depends on a variety of chamber-specific factors and, therefore, concentrations of DO measured in unimportant test chambers cannot be used.

Enter 20N if the most relevant information available is a statement that the concentration of DO in one or more important test chambers was below the lower limit on H given above for the test organisms at any time during the test.

If the instructions given above did not result in an entry, enter 30U if quantitative information is not available concerning the results of measurements of the concentration of DO in at least half of the important test chambers at least every other day. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. Enter a dash if all of the concentrations of DO that were measured in important test chambers were above the lower limit on H given above for the test organisms.

2. If sufficient information is available that both m and H can be derived from all of the concentrations of DO that were measured in important test chambers, the entry is assigned as follows, where:
   m = the arithmetic mean of the measured concentrations of DO.
   LLm = the lower limit on m given above for the test organisms.
   H = m - 1.96SD.
   LLH = the lower limit on H given above for the test organisms.
   x = the number of Ns:

   If $m \geq LLm$, enter a dash. If $H \geq LLH$, enter a dash.
   If $m < LLm$, $y = 40(LLm-m)$ If $H < LLH$, $z = 40(LLH-H)$

   $x = y + z$

   The maximum allowed value of x is 40. At least two decimal digits must be retained if x, y, and/or z are rounded.

   For salmonids, for example:
   If m = 6.4 and H = 4.9, enter 8.00N
   If m = 6.0 and H = 4.5, enter 40.00N

If both m and H cannot be derived as described above because the only results available are chamber-specific summaries of DO measurements, determine estimates of m and H as follows:

A. Calculate P = the arithmetic mean of the means for the individual test chambers and use P as an estimate of m.
B. Derive R = P-1.96SD for each test chamber and use the arithmetic mean of the Rs as an estimate of H.

The estimates of m and H are then used to determine the number of Ns, as described above.

If m can be derived or estimated, but H cannot be derived or estimated, the entry is assigned as follows:

1. If any Ns would be assigned because of m based on the above, enter twice the number of Ns.
2. If a dash would be assigned because of m based on the above, enter 10U.

Do not enter M.

Item E9

If information is available concerning the concentration of dissolved nitrogen in the test chambers, use Section A. Otherwise, use Section B.

Section A: Dissolved Nitrogen

Unless the purpose of the test required a high concentration of dissolved nitrogen (DN) in one or more test chambers, it seems reasonable that, in a high quality test, the concentration of DN in important test chambers should be $\leq$103% of saturation.

Enter a dash if the purpose of the test required that the concentration of DN be above 103% of saturation in one or more test chambers.

Unless there was substantial aeration or substantial change in water temperature, the concentration of DN in test solutions in test chambers is expected to be equal to or lower than the concentration of DN in the dilution water or in the test solutions before they entered test chambers. Therefore, in the absence of substantial aeration or change in water temperature, a measured concentration of DN in dilution water and/or test solutions before they entered the test chambers can be used as an upper limit on the concentration of DN in the test chambers during the test.

If the most relevant information available concerns the concentration of DN in the dilution water and/or test solutions before they entered the test chambers:

1. Enter a dash if the concentration of DN was below 103% of saturation.
2. Otherwise, enter 30U.

The concentration of DN that occurs in test chambers depends on a variety of chamber-specific factors and, therefore, concentrations of DN measured in unimportant test chambers cannot be used.

Enter 10U if the most relevant information available is a statement that all of the concentrations of DN that were measured in important test chambers were ≤103% of saturation.

Enter 30N if the most relevant information available is a statement that the mean of the concentrations of DN that were measured in important test chambers was > 103% of saturation.

Enter 15N if the most relevant information available is that the dilution water or test solutions were heated during the last 24 hours prior to entering the test chambers except enter a dash if the dilution water or test solutions were treated to remove supersaturated gases between the time they were heated and the time they entered the test chambers.

If the instructions given above did not result in an entry, enter 30U if quantitative information is not available concerning the results of measurements of the concentration of DN in at least two or more important test chambers at two or more times, including near the beginning and end of the test. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. Enter a dash if all of the concentrations of DN that were measured in important test chambers were ≤ 103% of saturation.

2. If each measured concentration of DN must be less than 103% of saturation, the chances of exceeding 103% increase as the number of measurements increases. To make the chances of exceedence independent of the number of measurements, the upper limit of 103% of saturation is applied to J = m+1.96SD, where m = the arithmetic mean of all of the concentrations of DN (as percent of saturation) that were measured in important test chambers and the standard deviation (SD) is derived from the same measured concentrations using the guidance presented in General Instruction #7.

   Enter a dash if one or more of the following are true:
   a. It is very unlikely that the dilution water was heated during the last 24 hours prior to entering the test chambers.
   b. The dilution water was treated to remove supersaturated gases during the last 24 hours prior to entering the test chambers.

   If the percent of saturation is calculated, it must be based on the mean test temperature, unless the temperature is known for each measured concentration of DN.

If sufficient information is available that J can be derived from all of the concentrations of DN that were measured in important test chambers, the entry is assigned as follows, where x = the number of Ns:

If J $\leq$ 103% of saturation, enter a dash.
If J > 103% of saturation, $x = 0.2(J-103)^2$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:

If J = 104%, enter 0.20N.
If J = 108%, enter 5.00N.
If J = 115%, enter 28.80N.

If J cannot be derived as described above because the only results available are treatment-specific or chamber-specific summaries of DN measurements, derive V = m+1.96SD for each test chamber and use the arithmetic mean of the Vs as an estimate of J, which is then used to determine the number of Ns, as described above.

If the mean measured concentration of DN can be derived or estimated, but J cannot be derived or estimated, enter a dash if the mean is $\leq$98% of saturation, but enter 30U if the mean is >98% of saturation.

Section B: Dissolved Oxygen

If sufficient measurements of the concentration of dissolved nitrogen (DN) are not available, results of measurements of dissolved oxygen (DO) should be used as surrogate information to determine the entry. Unless the purpose of the test required a high concentration of DO in one or more test chambers, it seems reasonable that, in a high quality test, the concentration of DO in important test chambers should be $\leq$ 100% of saturation. Because test organisms and BOD utilize oxygen, when the concentration of DO is above 100% of saturation, it is likely that the concentration of DN is higher than 103% of saturation.

Enter a dash if the purpose of the test required that the concentration of DO be above 100% of saturation in one or more test chambers.

Unless there was substantial aeration or substantial change in water temperature, the concentration of DO in test solutions in test chambers is expected to be equal to or lower than the concentration of DO in the dilution water or in the test solutions before they entered test chambers. Therefore, in the absence of substantial aeration or change in water temperature, a measured concentration of DO in dilution water and/or test solutions before they entered the test chambers can be used as an upper limit on the concentration of DO in the test chambers during the test.

If the most relevant information available concerns the concentration of DO in the dilution water and/or test solutions before they entered the test chambers:

1. Enter a dash if all of the measured concentrations of DO in the dilution water and/or test solutions before they entered the test chambers was below 100% of saturation.
2. Otherwise, enter 30U.

The concentration of DO that occurs in test chambers depends on a variety of chamber-specific factors and, therefore, concentrations of DO measured in unimportant test chambers cannot be used.

Enter 10U if the most relevant information available is a statement that all of the concentrations of DO that were measured in important test chambers were $\leq$100% of saturation.

Enter 30N if the most relevant information available is a statement that the mean of the concentrations of DO that were measured in important test chambers was > 100% of saturation.

Enter 15N if the most relevant information available is that the dilution water or test solutions were heated during the last 24 hours prior to entering the test chambers except enter a dash if the dilution water or test solutions were treated to remove supersaturated gases between the time they were heated and the time they entered the test chambers.

If the instructions given above did not result in an entry, enter 30U if quantitative information is not available concerning the results of measurements of the concentration of DO in at least two or more important test chambers at two or more times, including near the beginning and end of the test. If quantitative information is available concerning the specified data, there are two approaches that can justify assigning a dash. If either approach justifies a dash, a dash is assigned, even if the other approach does not justify a dash. The second approach will also assign Ns when appropriate.

1. Enter a dash if all of the concentrations of DO that were measured in important test chambers were $\leq$ 100% of saturation.

2. If each measured concentration of DO must be less than 100% of saturation, the chances of exceeding 100% increase as the number of measurements increases. To make the chances of exceedence independent of the number of measurements, the upper limit of 100% of saturation is applied to K = m+1.96SD, where m = the arithmetic mean of all of the concentrations of DO (as percent of saturation) that were measured in important test chambers and the standard deviation (SD) is derived from the same measured concentrations using the guidance presented in General Instruction #7.

Enter a dash if one or more of the following are true:

a. It is very unlikely that the dilution water was heated during the last 24 hours prior to entering the test chambers.
b. The dilution water was treated to remove supersaturated gases during the last 24 hours prior to entering the test chambers.

If the percent of saturation is calculated, it must be based on the mean test temperature, unless the temperature is known for each measured concentration of DO.

If sufficient information is available that K can be derived from all of the concentrations of DO that were measured in important test chambers, the entry is assigned as follows, where x = the number of Ns:

If $K \leq 100\%$ of saturation, enter a dash.
If $K > 100\%$ of saturation, $x = 0.2(K-100)^2$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:

If K = 101%, enter 0.20N.
If K = 105%, enter 5.00N.
If K = 112%, enter 28.80N.

If K cannot be derived as described above because the only results available are treatment-specific or chamber-specific summaries of DO measurements, derive W = m+1.96SD for each test chamber and use the arithmetic mean of the Ws as an estimate of K, which is then used to determine the number of Ns, as described above.

If the mean measured concentration of DO can be derived or estimated, but K cannot be derived or estimated, enter a dash if the mean is $\leq$95% of saturation, but enter 30U if the mean is >95% of saturation.

Do not enter M.

F. Test Results

Item F1

Toxicity that occurs during an acute exposure is “concurrent acute toxicity”, whereas toxicity that occurs shortly after an acute exposure is “delayed acute toxicity”. (Delayed acute toxicity is studied by placing the test organisms in dilution water with no added test material at the end of an acute exposure.) Both kinds of acute toxicity are important for evaluating the total severe acute toxicity of a test material. Although a toxicity test that provides information concerning both concurrent and delayed acute toxicity is more useful, a comparable test that only provides

information concerning concurrent toxicity or delayed toxicity is also useful. Further, a test result can often be useful even if it is not clear whether the result is based on only concurrent acute toxicity or on both concurrent and delayed acute toxicity. But these are issues of utility, not quality.

Enter 10U if there is an explicit reason for uncertainty as to whether the results are based on concurrent effects, delayed effects, or both.

Otherwise, enter a dash.

Item F2

See Item 14 on the Acute Suitability Checklist and its instructions.

Enter a dash if the test organisms were not worms.

Enter a dash if the test organisms were worms and reproduction did not occur in any test chambers.

If reproduction occurred during a test with worms, any "new" worms would have been produced by fission, which means that both the "new" and "old" worms were exposed from the beginning of the test. Thus, if the test organisms were worms and they reproduced during the test, the test result can be useful, but the calculation of the endpoint should appropriately account for reproduction.

The entry is determined as follows:

a. Enter a dash if, every day, the live organisms were counted, the dead organisms were counted and removed, and the result was based on the total number of worms that was in the chamber during the test and the total number of worms that died in the chamber during the test.
b. Enter 5N if, at the end of the test, the live organisms were counted, the dead organisms were counted, and the result was based on the numbers of live and dead organisms in the test chambers at the end of the test.
c. Enter 5N if, every day, the dead organisms were counted and removed, and the result was based on the total number of dead organisms in the test chambers and the number of organisms in the chambers at the beginning of the test. (If the total number of dead organisms in a test chamber is greater than the number in the chamber at the beginning of the test, the percent dead is assumed to be 100%.)
d. Enter 5N if, every day, the live organisms were counted, and the result was based on the total number of live organisms in the test chambers at the end of the test and the number of organisms in the chambers at the beginning of the test. (If the number of live organisms in a test chamber at the end of the test is greater than the number in the chamber at the beginning of the test, the percent live is assumed to be 100%.)
e. Enter 10N if, at the end of the test, the dead organisms were counted, and the result was based on the number of dead organisms in the test chambers at the

end of the test and the number of organisms in the test chambers at the beginning of the test.

It is possible that fission might affect the sensitivities of the organisms, but specially designed tests would have to be conducted to determine whether such an effect occurs.

Item F3

Death is considered to be a sign of disease, injury, and/or stress.

If there are two or more control treatments, (e.g., a solvent control and a dilution-water control), they are different control treatments and therefore the raw data are not pooled. If there was no control treatment, this item is applied to the lowest tested concentration of the test material. Replicate test chambers within a treatment must be pooled to determine the percent for the treatment. The entry is assigned as follows, where P = the highest percent of the organisms that exhibited signs of disease, injury, or stress in a control treatment (or in the lowest concentration of test material if there was no control treatment) and x = the number of Ns:

If $P \leq 15\%$, enter a dash.
If $P > 15\%$, $x = 2(P-15)$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:
If P = 16%, enter 2.00N.
If P = 25%, enter 20.00N

If it is reported that “some” or “a few” organisms were diseased, injured, or stressed, enter 10N.

The assigned entry is based on the amount of uncertainty that the various percentages cause in the results of the test; the assigned entry is not based on what is and is not considered feasible for various species.

If none of the above apply, enter 20U.

Do not enter M.

Item F4

If there are two or more control treatments (e.g., a solvent control and a dilution-water control), they are different control treatments and therefore the raw data are not pooled. If there was no control treatment, this item is applied to the lowest tested concentration of the test material. Replicate test chambers within a treatment must be pooled to determine the percent for the treatment. The entry is assigned as follows, where P = the highest percent of the organisms that died (if death is the

only effect being studied) or were affected (if one or more effects in addition to death are being studied) in a control treatment (or in the lowest concentration of test material if there was no control treatment) and x = the number of Ns:

If $P \leq 5\%$, enter a dash.
If $P > 5\%$, $x = 0.1(P-5)^2$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:
If P = 6%, enter 0.10N.
If P = 10%, enter 2.50N.
If P = 20%, enter 22.50N.

The assigned entry is based on the amount of uncertainty that the various percentages cause in the results of the test; the assigned entry is not based on what is and is not considered feasible for various species.

If none of the above apply, enter 20U.

Do not enter M.

Item F5

The confidence in a calculated value of an endpoint depends more on the number of test organisms exposed to concentrations that are close to the endpoint than on individual aspects of the experimental design such as the actual test concentrations, the dilution factor, and the number of organisms per test chamber and per treatment. If "important treatment" is defined as in the General Instructions and if tests are conducted with 20 organisms per treatment at a dilution factor of 0.6, many acute tests will have 40 organisms exposed in important treatments.

If v (= the total number of test organisms exposed in important treatments) is unknown, enter 20U.

If v is known, the entry is assigned as follows, where v = the total number of test organisms exposed in important treatments and x = the number of Ns:

If $v \geq 40$, enter a dash.
If $v < 40$, $x = [(0.4)(40-v)][1+(40/v)(0.18)]$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:
If v = 30, enter 4.96N.
If v = 20, enter 10.88N.
If v = 10, enter 20.64N.

If v = 4, enter 40.00N.

As stated in the General Instructions, although important treatments provide the most useful information, other treatments can provide confirmatory information. In theory, it would be possible to take into account how close each treatment is to the endpoint and the number of test organisms in each treatment, but this is considered an unnecessary level of detail.

Do not enter M.

Item F6

Replicate test chambers must be pooled to determine the percent for the treatment.

P is calculated for a treatment using Abbott's formula (Finney 1971):

$P = (P^*-C)/(100-C)$

where:

P = the adjusted percent for the treatment.
P* = the observed percent for the treatment.
C = the observed or estimated percent for the control treatment.

If an LC50 or EC50 is stored in EVISTRA, the entry is assigned as follows, where P = the adjusted percent of organisms that died or were affected in the lowest concentration that was not a control treatment and x = the number of Ns:

If $P \leq 37\%$, enter a dash.
If $P > 37\%$, $x = P-37$.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:
If P = 38%, enter 1.00N.
If P = 50%, enter 13.00N.
If P = 60%, enter 23.00N.

If the lowest tested concentration is stored in EVISTRA as a "less than" value, enter a dash even if the percent of organisms that died or were affected in the lowest tested concentration was more than 50%. It might be better to use a "less than" value that has a higher rating than a calculated point estimate of an endpoint that has a lower rating.

If none of the above apply, enter 20U.

Do not enter M.

Item F7

Replicate test chambers must be pooled to determine the percent for the treatment.

P is calculated for a treatment using Abbott's formula (Finney 1971):

P = (P*-C)/(100-C)

where:

P = the adjusted percent for the treatment.

P* = the observed percent for the treatment.

C = the observed or estimated percent for the control treatment.

If an LC50 or EC50 is stored in EVISTRA, the entry is assigned as follows, where P = the adjusted percent of organisms that died or were affected in the highest concentration and x = the number of Ns:

If P $\geq$ 63%, enter a dash.

If P < 63%, x = 63-P.

The maximum allowed value of x is 40. At least two decimal digits must be retained if x is rounded.

For example:

If P = 62%, enter 1.00N.

If P = 50%, enter 13.00N.

If P = 40%, enter 23.00N.

If the highest tested concentration is stored in EVISTRA as a "greater than" value, enter a dash even if the percent of organisms that died or were affected in the highest tested concentration was less than 50%. It might be better to use a "greater than" value that has a higher rating than a calculated point estimate of an endpoint that has a lower rating.

If none of the above apply, enter 20U.

Do not enter M.

Item F8

Enter N if available data indicate that there was a substantial discontinuity in the slope.

If EC50s, etc., are available for only two durations (and the first is at least 20 hours and the second is at least 1.9 times the first), enter a dash if the EC50 for the longer time is equal to or up to five times lower than the EC50 for the shorter time. Also enter a dash if the same or different investigators obtained a comparable slope in another test on the same test material with the same or a different test organism. (This might occasionally require a consideration of slopes obtained in other tests.)

Enter U if only one LC50, etc., is available or if no relevant information is available.

Do not enter M.

G. Measurement of Concentrations of Test Material in Test Solutions

The score for Section G depends on whether the concentrations of the test material were measured in test solutions, various aspects of the measurement methodology, properties of the test material, and various aspects of the toxicity test methodology. The relevant issues are addressed in Subsections G1, G2, and G3 below. The approach used in Section G is that Subsection G1 provides an interim score, and then Subsection G2 provides a multiplier of the interim score to take into account situations in which it is especially important to measure the concentration of the test material, i.e., situations in which unmeasured concentrations are especially questionable (see item h on page 6 and the examples on page 11). Subsection G3 addresses a related issue.

Subsection G1

If the concentration of test material was not measured in any test solution, enter 2N for items G1a, G1b, G1c, and G1d; in addition, enter 1N for G1e and G1f and then go to the section of the instructions titled “Result for Subsection G1".

If the concentration of test material was measured, follow the instructions below for items G1a through G1f to determine the entry for Subsection G1.

Item G1a

In order for the analytical method to have been sufficiently selective, it must have been able to differentiate the test material from likely degradation products and impurities and from other substances in the dilution water. For example, a colorimetric method that detects a variety of phenols is not sufficiently selective because it will measure all impurities that are phenols. As necessary, the EVISTRA coordinator will provide detailed guidance concerning analytical methods for measuring test materials for which results are evaluated for EVISTRA.

If the analytical method was sufficiently selective, enter a dash.

If the analytical method was not sufficiently selective, enter 2N.

If unknown, enter 2U.
Do not enter M.

Item G1b

In order for the samples to have been appropriate, the following requirements must have been satisfied:

1. For static tests, the concentration of test material must have been determined in samples taken from each important treatment at least at both the beginning and end of the test.
2. For renewal tests, the concentration of test material must have been determined in samples taken (a) from each important treatment at both the beginning and end of at least one renewal and (b) during at least two renewals.

3. For flow-through tests, the concentration of test material must have been determined in samples taken from each important treatment at least near the beginning and end of the test.

Enter a dash if the analyzed samples of test solutions satisfied the relevant requirement specified above.

Enter 0.02(100-P)N if the analyzed samples of test solutions satisfied about P percent of the relevant requirements specified above.

Enter 2N if the analyzed samples satisfied none of the relevant requirements specified above.

If unknown, enter 2U.
Do not enter M.

Item G1c

Samples of test solutions must have been handled (obtained, stored, shipped, etc.) in ways that prevented loss of test material by sorption, volatilization, photodegradation, biodegradation, precipitation, etc. The best way to demonstrate that there probably was no loss is to handle spiked samples and/or spiked control water along with the samples.

Enter a dash if (1) spiked samples and/or spiked control water showed negligible loss of test material, or (2) it appeared that adequate measures were taken to ensure that loss of test material did not occur during handling of the samples.

Enter 1N if it appears that there could have been some loss of test material during handling of the samples.

Enter 2N if it appears likely or quite possible that there was substantial loss of test material during handling of the samples.

If unknown, enter 2U.
Do not enter M.

Item G1d

If the percent recovery was not reported, but the raw data are available, for each sample that was analyzed, spiked, and reanalyzed, calculate the percent recovery using the following equation:

$$R = 100(A-B)/C$$

where:

R = the percent recovery.
A = the concentration measured in the reanalysis after the spike was

added.

B = the concentration measured in the initial analysis.

C = the concentration spiked into the sample.

Then calculate the arithmetic mean of the Rs.

Enter a dash if at least ten percent of the analyzed samples of stock solution and test solution were spiked with a known quantity of test material and reanalyzed and if the mean recovery was ≥ 87% and ≤ 115%.

Otherwise:

1. Enter 2N if no samples were spiked and reanalyzed.
2. Enter 1.33N if both the percent of the samples that were spiked and reanalyzed was > 0% and < 10% and the mean recovery was < 87% or > 115%.
3. Enter 0.67N if either the percent of the samples that were spiked and reanalyzed was > 0% and < 10% or the mean recovery was < 87% or > 115%.

If unknown, enter 2U.

Do not enter M.

Item G1e

For each sample that was analyzed twice, divide the higher measured concentration by the lower measured concentration to obtain a high-low quotient. Then calculate the geometric mean of all of the quotients.

Enter a dash if at least ten percent of the analyzed samples of stock solution and test solution were analyzed twice and if the geometric mean of all of the quotients was ≤ 1.15.

Otherwise:

1. Enter 1N if there were no replicate analyses.
2. Enter 0.67N if both the percent of the samples that were analyzed twice was > 0% and < 10% and the geometric mean of all of the quotients was > 1.15.
3. Enter 0.33N if either the percent of the samples that were analyzed twice was > 0% and < 10% or if the geometric mean of all of the quotients was > 1.15.

If unknown, enter 1U.

Do not enter M.

Item G1f

An acceptable reference standard is a standard solution that is obtained from a company or organization that prepares such standards. A reference standard cannot be made in the same laboratory by weighing a material and dissolving it in a solvent. The reference standard may have been used to

prepare test solutions and may have been diluted in order to prepare a standard curve.

A reference standard was "used" even if it was only used to verify the concentrations in standards prepared in the laboratory.

If an acceptable reference standard was not used, enter 1N.
If there is no mention of a reference standard, enter 1U.
Do not enter M.

Result for Subsection G1:

If all of the entries for items G1a through G1f are dashes, enter a dash for G1, skip Subsection G2, and go to Subsection G3.

If the concentration of test material was not measured in one or more test solutions, enter G1=10N; then go to Subsection G2.

If the concentration of test material was measured, but one or more Ns were assigned to items G1a through G1f, sum the number of Ns assigned to items G1a through G1f and enter the total (be sure to include the N) as G1; then go to Subsection G2.

If the concentration of test material was measured and no Ns were assigned to items G1a through G1f but one or more Us were assigned to items G1a through G1f, sum the number of Us assigned to items G1a through G1f and enter the total (be sure to include the U) as G1; then go to Subsection G2.

Subsection G2

As per above, skip Subsection G2 if the entry for G1 is a dash.

Item G2a. Volatility

1. For tests that EPA wants evaluated for EVISTRA, the EVISTRA coordinator will provide an estimate of the number of hours it takes for 50% of the test material to volatilize from still water (i.e., an estimate of the half-life) at 25°C. If the estimate is 1000 or more hours, enter A=1000 and enter G2a=0; then go to Item G2b.
2. The "average residence time" is defined as:
   a. For static tests - the duration of the test (in hours).
   b. For renewal tests - the average number of hours that the test organisms were in the renewals (e.g., if the test solution was renewed every 24 hours, the average residence time was 24 hours.)
   c. For flow-through tests - the volume of solution in the test chamber divided by the average flow rate (volume per hour) of test solution through the test chamber. (This approximation does not take the degree of mixing into account because making measurements or

making an assumption concerning the degree of mixing is considered an unnecessary level of detail.)

The average residence time must be expressed in hours.

The half-life is used to estimate the percent of the test material that is lost during the average residence time, using the equations for a first-order reaction (Moore 1972):

A = half life (in hours)
B = average residence time (in hours)
X = percent of test material remaining after B hours
$= (100)(e^{-[\ln 2][B/A]})$
Y = percent lost during B hours
= 100 - X

The weighting is then a function of Y = percent lost. If G1 = 10N because the concentration of test material was not measured acceptably in any test solutions and if it is estimated that 50% of the test material would be lost during the average residence time, it seems reasonable that the weight should be about 4 so that the test would receive at least 40N for G and its quality rating would be no higher than “very low”. In contrast, if G1 = 10N and if it is estimated that 5% of the test material would be lost during the average residence time, it seems reasonable that the weight should be about 1 so that the test would receive at least 10N for G and its quality rating would be no better than “high”. Therefore, it seems reasonable to calculate the weight using the following equation:

Z = weight
$= (0.2Y)^{0.6}$

For example:

A (Half-life) (hours)	B (Ave. Ret. Time) (hours)	Y (% Lost)	Z (Weight)
24	0.5	1.4	0.47
24	1.8	5.1	1.01
24	24	50.0	3.98
24	96	93.8	5.80
24	192	99.6	6.02

Because of the usual difference in average residence times, the fraction of a volatile test material that volatilizes will usually be lower in a flow-through test than in a comparable static test and so the flow-through test

will receive a higher rating than a static test, if there are no other differences between the tests.

3. The effect of volatility needs to be modified if the test temperature is different from the 25°C on which the estimate of half-life is based. Because volatility approximately doubles with a ten C degree increase in temperature, 1.414 (i.e., the square root of 2) is raised to a power that equals the difference in temperature divided by 5 so that a difference of ten C degrees results in a factor of 2 change in the weight.

   The "test temperature" should be determined as the arithmetic mean of all of the temperatures measured in important test chambers, if possible, or as the arithmetic mean of all of the temperatures measured in all of the test chambers. If neither of these is possible, the "test temperature" should be determined as the arithmetic mean of the mean measured temperatures for each of the important treatments, if possible, or as the arithmetic mean of the mean measured temperatures for all of the treatments. If none of the above are possible, the "test temperature" should be determined as the midpoint of the range of the temperatures measured in important test chambers, if possible, or as the midpoint of the range of the temperatures measured in all of the test chambers.
4. Because aeration will increase volatilization, an arbitrary value of 2 is used to account for the increased loss of test material during aerated tests. This value of 2 may be modified by the EVISTRA coordinator if data are available to allow derivation of a better estimate of the effect of aeration on volatilization of the test material. The effect of aeration is considered multiplicative because volatilization from the test solutions will occur even if the solutions are not aerated; aeration merely increases the rate of volatilization.
5. If test solutions of a volatile test material are prepared by evaporating stock solution onto a surface of the test chambers and adding water, more test material is likely to be lost during preparation of the test solutions than during the test itself. This effect of this procedure is considered to be additive because it is distinct from volatilization from test solutions.

Item G2b. Hydrolysis

1. For tests that EPA wants evaluated for EVISTRA, the EVISTRA coordinator will provide an estimate of the number of hours it takes for 50% of the test material to hydrolyze (i.e., an estimate of the half-life) at 25°C. If the estimate is 1000 or more hours, enter F=1000 and enter G2b=0; then go to Item G2c.
2. The "average residence time" is defined as:
   a. For static tests - the duration of the test (in hours).
   b. For renewal tests - the average number of hours that the test organisms were in the renewals (e.g., if the test solution was renewed every 24 hours, the average residence time was 24 hours.)

c. For flow-through tests - the volume of solution in the test chamber divided by the average flow rate (volume per hour) of test solution through the test chamber. (This approximation does not take the degree of mixing into account because making measurements or making an assumption concerning the degree of mixing is considered an unnecessary level of detail.)

The average residence time must be expressed in hours.

The half-life is used to estimate the percent of the test material that is lost during the average residence time, as explained in Item G2a.

3. The effect of hydrolysis needs to be modified if the test temperature is different from the 25°C on which the estimate of half-life is based. Because hydrolysis increases with temperature at about 10% per C degree, the entry for hydrolysis is modified by multiplying by 1.1 to the power of the difference in temperature. This results in a 10% increase in the weight for every C degree that the test temperature is above 25°C and a 10% decrease in the weight for every C degree that the test temperature is below 25°C.

   The "test temperature" should be determined as the arithmetic mean of all of the temperatures measured in important test chambers, if possible, or as the arithmetic mean of all of the temperatures measured in all of the test chambers. If neither of these is possible, the "test temperature" should be determined as the arithmetic mean of the mean measured temperatures for each of the important treatments, if possible, or as the arithmetic mean of the mean measured temperatures for all of the treatments. If none of the above are possible, the "test temperature" should be determined as the midpoint of the range of the temperatures measured in important test chambers, if possible, or as the midpoint of the range of the temperatures measured in all of the test chambers.

The rate of hydrolysis can also be affected by pH, but the effect of pH varies substantially from one chemical to another and is not important for most chemicals when pH is near 7. Therefore, the effect of pH on the amount of loss due to hydrolysis is not routinely taken into account here.

Item G2c. Biodegradation

1. For tests that EPA wants evaluated for EVISTRA, the EVISTRA coordinator will provide an estimate of the number of hours it takes for 50% of the test material to biodegrade (i.e., an estimate of the half-life) at 25°C. If the estimate is 1000 or more hours, enter J=1000 and enter G2c=0; then go to Item G2d.
2. The "average residence time" is defined as:
   a. For static tests - the duration of the test (in hours).

b. For renewal tests - the average number of hours that the test organisms were in the renewals (e.g., if the test solution was renewed every 24 hours, the average residence time was 24 hours.)
c. For flow-through tests - the volume of solution in the test chamber divided by the average flow rate (volume per hour) of test solution through the test chamber. (This approximation does not take the degree of mixing into account because making measurements or making an assumption concerning the degree of mixing is considered an unnecessary level of detail.)

The average residence time must be expressed in hours.

The half-life is used to estimate the percent of the test material that is lost during the average residence time, as explained in Item G2a.

3. Because aeration will generally increase the rate of biodegradation, an arbitrary value of 2 is used to account for the increased loss of test material during aerated tests. This value of 2 may be modified by the EVISTRA coordinator if data are available to allow derivation of a better estimate of the effect of aeration on biodegradation of the test material. The effect of aeration is considered multiplicative because biodegradation will occur in the test solutions even if the solutions are not aerated; aeration merely increases the rate of biodegradation.
4. Because the presence of elevated amounts of organic carbon will generally increase the rate of biodegradation, an arbitrary value of 2 is used to account for the increased loss of test material when the test chambers contain elevated amounts of organic carbon. The presence of a substrate is considered to be a multiplicative effect because its major impact is to increase the concentration of organic carbon in the test solutions.

The effect of temperature on the rate of biodegradation is so chemical-specific that it is not taken into account here.

Item G2d. Sorption of organic chemicals

1. Log = logarithm to base 10.

The tendency to sorb to organic carbon (in excretion products, substrate, and food) and surfaces increases as log Kow increases so the entry is N=log Kow. If log Kow is less than 4, however, it is unlikely that sorption will be a problem; therefore, if log Kow<4, enter G2d=0; then go to Item G2e.

2. The “average residence time” is defined as:
   a. For static tests - the duration of the test (in hours).
   b. For renewal tests - the average number of hours that the test organisms were in the renewals (e.g., if the test solution was renewed every 24 hours, the average residence time was 24 hours.)

c. For flow-through tests - the volume of solution in the test chamber divided by the average flow rate (volume per hour) of test solution through the test chamber. (This approximation does not take the degree of mixing into account because making measurements or making an assumption concerning the degree of mixing is considered an unnecessary level of detail.)

The average residence time must be expressed in hours.

The amount of sorption will increase as the average residence time increases. If G1 = 10N because no measurements were made on test solutions, it seems reasonable that even a flow-through test with a chemical whose log Kow = 6 should receive a quality rating no higher than "very low".

3. Silica sand does not contain organic carbon, but most other materials used as substrates do. For tests that EPA wants evaluated for EVISTRA, the EVISTRA coordinator will provide information as necessary concerning substrates other than silica sand. The effect of a substrate is considered to be additive because the four causes of sorption are distinct. An arbitrary value of 1 is used because the four causes are considered equally important.
4. All foods for aquatic organisms contain organic carbon. The effect of food is considered to be additive because the four causes of sorption are distinct. An arbitrary value of 1 is used because the four causes are considered equally important.
5. For tests that EPA wants evaluated for EVISTRA, the EVISTRA coordinator will provide information as necessary concerning other materials. The effect of construction materials is considered to be additive because the four causes of sorption are distinct. An arbitrary value of 1 is used because the four causes are considered equally important.

Item G2e. Conditions that can affect the toxicities of cations of heavy metals

1. Cations of such heavy metals as cadmium, copper, lead, nickel, silver, and zinc interact with organic carbon and sorb to some surfaces.
2. The "average residence time" is defined as:
   a. For static tests - the duration of the test (in hours).
   b. For renewal tests - the average number of hours that the test organisms were in the renewals (e.g., if the test solution was renewed every 24 hours, the average residence time was 24 hours.)
   c. For flow-through tests - the volume of solution in the test chamber divided by the average flow rate (volume per hour) of test solution through the test chamber. (This approximation does not take the degree of mixing into account because making measurements or making an assumption concerning the degree of mixing is considered an unnecessary level of detail.)

The average residence time must be expressed in hours.

The average residence time is used to weight the factors because the reactions are not instantaneous. If G1 = 10N because no measurements were made on test solutions, it seems reasonable that even a flow-through test with a cation of a heavy metal should receive a quality rating no higher than "high" if even one of the relevant special conditions occurred.

3. Silica sand does not contain organic carbon, but most other materials used as substrates do. For tests that EPA wants evaluated for EVISTRA, the EVISTRA coordinator will provide information as necessary concerning substrates other than silica sand. The effect of a substrate is considered to be additive because the four reactions of metal cations are distinct. An arbitrary value of 4 is used because the four reactions are considered to be equally important.
4. All foods for aquatic organisms contain organic carbon. The effect of food is considered to be additive because the four reactions of metal cations are distinct. An arbitrary value of 4 is used because the four reactions are considered to be equally important.
5. For tests that EPA wants evaluated for EVISTRA, the EVISTRA coordinator will provide information as necessary concerning other materials. The effect of test materials is considered to be additive because the four reactions of metal cations are distinct. An arbitrary value of 4 is used because the four reactions are considered to be equally important.
6. If test solutions of a cation of a heavy metal are prepared by evaporating stock solution onto a surface of the test chambers and adding water, it is unlikely that all of the test material will redissolve in the dilution water. The effect of this procedure is considered to be additive because it is distinct from the other reactions of cations of heavy metals. An arbitrary value of 4 is used because the four reactions are considered to be equally important.

Result for Subsection G2:

The entry for G2 cannot be less than 1 because a test cannot be high quality if the entry for G1 is 10N.

Subsection G3

Enter a dash if (a) the test result reported in the document was based on measured concentrations of the test material in the test solutions or (b) a new result can be calculated based on measured concentrations in test solutions so that the new result can be stored in EVISTRA.

Enter 4N if the reported result was not based on measured concentrations and sufficient information is not reported to allow calculation of a result based on measured concentrations in test solutions.

Enter 4U if it is not known whether the reported result was based on measured concentrations in test solutions and sufficient information is not reported to allow calculation of a result based on measured concentrations in test solutions. If the report does not say that the result was based on measured concentrations in test solutions, do not assume that it was, even the report says that the measured concentrations were corrected for recovery.

Do not enter M.

Entry for G:

Enter a dash if, and only if, the entries for both G1 and G3 are dashes.

If U or N is entered, the weight cannot be greater than 40.

H. Miscellaneous

It is not practical or possible for a checklist to cover everything that might affect the level of confidence in a test result. New issues will arise occasionally and some issues occur so rarely that they do not need to be addressed in a checklist. For example:

1. The concentration of Total Suspended Solids (TSS) might be sufficiently high in dilution water that it will stress the test organisms.
2. Even though raw data are not available, it might be reported that:
   a. There was an inversion or discontinuity in the concentration-effect curve.
   b. The data for % effect vs. time cannot be fit reasonably well with a smooth curve for each test chamber and treatment.

If an unusual situation arises when a test result is being evaluated for EPA, the EVISTRA coordinator must be consulted.

Do not enter U or M.

References

ASTM. 1998a. Standard Guide for Conducting Static Acute Toxicity Tests Starting with Embryos of Four Species of Saltwater Bivalve Molluscs. Standard E724. Annual Book of ASTM Standards, Vol. 11.05. West Conshohocken, PA.

ASTM. 1998b. Standard Guide for Conducting Acute Toxicity Tests on Test Materials with Fishes, Macroinvertebrates, and Amphibians. Standard E729. Annual Book of ASTM Standards, Vol. 11.05. West Conshohocken, PA.

Bosnak, A.D., and E.L. Morgan. 1981. Acute Toxicity of Cadmium, Zinc, and Total Residual Chlorine to Epigean and Hypogean Isopods (Asellidae). Natl. Speleological Soc. Bulletin 43:12-18.

Finkel, A.M. 1990. Confronting Uncertainty in Risk Management. Center for Risk Management, Washington, DC.

Finney, D.J. 1971. Probit Analysis. 3rd ed. Cambridge University Press, London. pp. 125-128.

Johnson, W.W., and M.T. Finley. 1980. Handbook of Acute Toxicity of Chemicals to Fish and Aquatic Invertebrates. Resource Publication 137. U.S. Fish and Wildlife Service, Washington, CD.

Mayer, F.L., Jr. 1987. Acute Toxicity of Chemicals to Estuarine Organisms. EPA/600/8-87/017. National Technical Information Service, Springfield, VA.

Mayer, F.L., Jr., and M.R. Ellersieck. 1986. Manual of Acute Toxicity: Interpretation and Data Base for 410 Chemicals and 66 Species of Freshwater Animals. Resource Publication 160. U.S. Fish and Wildlife Service, Washington, DC.

McNulty, E.W., F.J. Dwyer, M.R. Ellersieck, E.I. Greer, C.G. Ingersoll, and C.F. Rabeni. 1999. Evaluation of Ability of Reference Toxicity Tests to Identify Stress in Laboratory Populations of the Amphipod *Hyalella azteca*. Environ. Toxicol. Chem. 18:544-548.

Merritt, R.W., and K.W. Cummins. 1996. An Introduction to the Aquatic Insects of North America. Third edition. Kendall/Hunt Publishing Co., Dubuque, IA.

Moore, W.J. 1972. Physical Chemistry. 4th ed. Prentice-Hall, Inc., Englewood Cliffs, NJ. pp. 333-334.

Pennak, R.W. 1978. Fresh-Water Invertebrates of the United States. 2nd ed. Wiley-Interscience. New York, NY. pp. 682-684.

Seiler, F.A., and J.L. Alvarez. 1998. On the Use of the Term *Uncertainty*. SETAC News. Society for Environmental Toxicology and Chemistry. November. pp. 23-24.

Snedecor, G.W., and W.G. Cochran. 1967. Statistical Methods. 6th ed. Iowa State University Press, Ames, IA. pp. 39, 40, and 47.

Sokal, R.R., and F.J. Rohlf. 1969. Biometry. Freeman and Co., San Francisco, CA.

Sprague, J.B. 1973. The ABCs of Pollutant Bioassay Using Fish. In: Biological Methods for the Assessment of Water Quality. ASTM STP 528. American Society for Testing and Materials, Philadelphia, PA. pp.6-30.

Swanson, M.B., and A.C. Socka, Eds. 1997. Chemical Ranking and Scoring. SETAC Press, Pensacola, FL.

U.S. EPA. 1985a. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses. PB85-227049. National Technical Information Service, Springfield, VA.

U.S. EPA. 1985b. Ambient Water Quality Criteria for Chlorine - 1984. EPA 440/5-84-030. National Technical Information Service, Springfield, VA.

U.S. EPA. 1986. Ambient Water Quality Criteria for Dissolved Oxygen. EPA 440/5-86-003. National Technical Information Service, Springfield, VA.